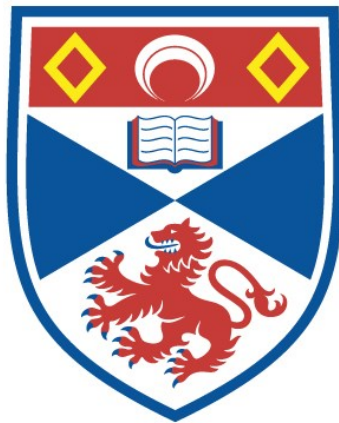


**WHY SHOULD I BE MORAL? : TOWARD A DEFENCE OF  
THE CATEGORICITY AND NORMATIVE AUTHORITY OF  
MORAL CONSIDERATIONS**

**Kent Hurtig**

**A Thesis Submitted for the Degree of PhD  
at the  
University of St Andrews**



**2004**

**Full metadata for this item is available in  
St Andrews Research Repository  
at:**

**<http://research-repository.st-andrews.ac.uk/>**

**Please use this identifier to cite or link to this item:**

**<http://hdl.handle.net/10023/13223>**

**This item is protected by original copyright**

# **WHY SHOULD I BE MORAL?**

Toward a Defence of the Categoricity and  
Normative Authority of Moral Considerations

by

Kent Hurtig

Dissertation submitted for the degree of Doctor of Philosophy

University of St Andrews

February, 2004





ProQuest Number: 10166937

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10166937

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code  
Microform Edition © ProQuest LLC.

ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 – 1346

TL  
E683

## ABSTRACT

Can we ever be fully practically justified in acting contrary to moral demands? My contention is that the answer is 'no'. I argue that by adopting a 'buck-passing' account of wrongness we can provide a philosophically satisfying answer to the familiar 'why should I be moral?'. In working my way toward the buck-passing account of wrongness, I outline (and, to some degree, defend) the metaethical and 'metanormative' assumptions on which my theory stands. I also consider and reject the 'internalist' (or as it can also be described, the neo-Humean) answer to 'why should I be moral?'. The account I end up with is decidedly non-consequentialist and it is consistent with common-sense morality. It also provides a way of showing why moral considerations (in competition with non-moral considerations) are overridingly normative in a way that is consistent with our best current understanding of what practical reason requires of us.

## DECLARATIONS

I, Kent Hurtig, hereby certify that this dissertation, which is approximately 64, 000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Date: 16/2/04

Signature of candidate:

I was admitted as a research student in January 1999 and as candidate for the degree of Doctor of Philosophy in January 1999; the higher study for which this is a record was carried out in the University of St Andrews between 1999 and 2004.

Date: 16/2/04

Signature of candidate:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St Andrews and that the candidate is qualified to submit this dissertation in application for that degree.

Date: 16/02/04

Signature of supervisor:

In submitting this dissertation to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any *bona fide* library or research worker.

Date: 16/2/04

Signature of candidate:

## ACKNOWLEDGEMENTS

Many people have been tremendously helpful to me while I have been writing this dissertation. I am especially thinking here of my supervisors Professor John Skorupski and Professor Sarah Broadie. I am grateful to both for their encouragement and patience. I am particularly grateful to John Skorupski for urging on me the importance of not getting too bogged down with 'interesting but irrelevant' questions that have no direct bearing on the issues under consideration. Whether or not, or to what degree, I have managed to take on board his sound advice is for the reader to decide. I am also indebted to Professor Broadie for her crucial encouragement and support during the latter stages of writing this dissertation. Both my supervisors' comments on my work have been very helpful indeed; again, I thank them both.

Thanks also to Dr David Archard and Dr Peter Clark for agreeing to act as referees. Special thanks to Dr Patrick Greenough for encouragement and for very useful advice on a number of issues - not all of which related to my dissertation, and to Simon Robertson who, I am sure, has read this dissertation from cover to cover several times. I have learnt a great deal from our frequent conversations and I wish you all the best with finishing your own dissertation. Other people who deserve thanks for many things -

for philosophical discussions as well as for, simply, being good friends – are: Iwao Hirose, Brian McElwee, and Raffaele Rodogno – congratulations to Iwao and Raffa for passing your examinations; and to Brian, good luck with your dissertation. Many thanks also to David and Fiona Dunleavey, Brett Langridge, Donald Burnett, and Magnus Sundström for being really good friends and for helping me out in various ways on numerous occasions. Very special thanks to Dr Despina Mavridou for steadfast and crucial encouragement during difficult times and for doing the right thing at the right time.

Finally, I owe a great debt of gratitude to my family for encouragement and support (not least of all financial): my mother Ingalill, my father Rolf, and my brother Anders. I love you all.

Kent Hurtig

Dunedin/St Andrews, 15 February, 2004

# TABLE OF CONTENTS

<b>INTRODUCTION.....</b>	<b>VII</b>
<b>CHAPTER ONE: A RATIONAL FOUNDATION FOR MORALITY? .....</b>	<b>12</b>
1.1 INTRODUCTION .....	12
1.2 'BEING MORAL' AND THE CHOICE OF METHODOLOGY .....	16
1.3 METAETHICS AND METANORMATIVITY.....	23
1.4 PRICHARD'S INTUITIONIST CHALLENGE (1) .....	31
1.5 PRICHARD'S INTUITIONIST CHALLENGE (2) .....	39
1.6 DIFFERENT CONCEPTIONS OF 'RATIONAL' .....	555
1.7 PRACTICAL REASON <i>VERSUS</i> PRACTICAL REASONS? .....	666
1.8 REASONS .....	733
1.9 SUMMARY AND CONCLUSION.....	81
<b>CHAPTER TWO: INTERNAL REASONS.....</b>	<b>87</b>
2.1 INTRODUCTION .....	87
2.2 DIFFERENT INTERNALISMS .....	91
2.3 WILLIAMS' INTERNALISM .....	93
2.4 DISAMBIGUATING INTERNALISM .....	99
2.5 TWO FURTHER DISTINCTIONS .....	104
2.6 POSSIBILISM AND MOTIVES .....	106
2.7 COUNTEREXAMPLES TO POSSIBILISM.....	111
2.8 ACTUALISM AND ACCIDIE.....	127
2.9 ACTUALISM AND REASONABLE REGRET .....	140
2.10 A FINAL ARGUMENT .....	145
2.11 SUMMARY AND CONCLUSION.....	153
<b>CHAPTER THREE: BELIEFS, DESIRES, AND MOTIVATION.....</b>	<b>158</b>
3.1 INTRODUCTION .....	158
3.2 THE ROLE OF BELIEFS IN INTERNALISM.....	160
3.3 THE CONTENT OF BASIC NORMATIVE BELIEFS .....	171
3.4 BASIC NORMATIVE BELIEFS AS TACIT BELIEFS .....	176
3.5 INTERNAL REASONS AND PROPOSITIONAL ATTITUDES HOLISM.....	181
3.6 DESIRE-BASED REASONS .....	190
3.7 DESIRES AND INSTRUMENTAL REASONS .....	198
3.8 BELIEFS AS SOURCES OF MOTIVATION .....	205
3.9 SUMMARY AND CONCLUSION.....	212
<b>CHAPTER FOUR: EXTERNALISM AND WHY WE SHOULD BE MORAL .....</b>	<b>216</b>
4.1 INTRODUCTION .....	216
4.2 EXTERNALISM, REASONS, AND VALUE .....	217
4.3 SCANLON'S DILEMMA.....	227
4.4 SCANLON'S CONTRACTUALISM AS AN ANSWER TO THE DILEMMA.....	233
4.5 A BUCK-PASSING ACCOUNT OF WRONGNESS? .....	240
4.6 SCANLON'S DILEMMA REVISITED .....	249
4.7 'REASON', 'A REASON', AND 'OUGHT' .....	256
4.8 BUCK-PASSING AND WHY WE OUGHT TO BE MORAL.....	265
4.9 PROBLEMS AND PROSPECTS FOR THE BUCK-PASSING ACCOUNT .....	272
<b>BIBLIOGRAPHY .....</b>	<b>287</b>

## INTRODUCTION

Most of us believe that some actions are right and that some actions are wrong; others believe that there are no such actions. Many of those who believe that there are right and wrong actions also hold that the wrongness of an action gives us good reason not to do it. Others, although they agree that there are such things as right and wrong actions, believe that an action's wrongness gives us little or no reason not to perform such actions. These people cannot all be right. In this dissertation I will address these issues in an attempt to answer one of the oldest questions in moral philosophy: why should I be moral?

'The Moral Question', as I will call it, is a question that can be approached from several different angles. The question may be, and frequently has been, asked as a request for a defence or elucidation of some metaphysical, epistemological, or normative issue thought to be problematic with or within moral discourse or practice. My main interest in this dissertation is not metaethical or metanormative (although I do have some things to say about these issues), with one exception: I am interested in the *justification* of morality. We are all familiar with the idea of theoretical, or epistemic, justification, but what could a justification of morality consist in? This is the main question I am trying to answer in this



dissertation. A sound working hypothesis is, I believe, that a justification of morality - if one is to be found - must be sought in a sound theory of practical reason. As I understand 'justification' in its practical sense, a justification of morality must show not only that moral considerations are genuinely normative, it must also show that these considerations are overridingly or trumpingly normative.

Although I am interested in investigating the possibility of there being a rational, or justificatory, foundation for morality, I am not particularly interested in what may be called the content of morality. A working hypothesis of mine is that common-sense morality is roughly right. Indeed, it might be true that insofar as any morality can be said to be right, common-sense morality must, *by conceptual necessity*, be at least roughly right. Moral functionalists, for example, will tell us that, in spite of Humpty Dumpty's declaration that we are entitled to mean what we like by our words,

...if we wish to address the concerns of our fellows when we discuss the matter - and if we don't, we will not have much of an audience - we had better mean what they mean. We had better, that is, identify our subject via the *folk* theory of rightness, wrongness, goodness, badness, and so on. We need not identify rightness as the property

that satisfies, or near enough satisfies, the folk theory of rightness - and likewise for the other moral properties. It is, thus, folk theory that will be our guide in identifying rightness, goodness, and so on.<sup>1</sup>

According to common-sense (or 'folk') morality, moral considerations are very much like Kant said they were: categorical and authoritative; i.e. they apply to all agents independently of their inclinations, and their normative force is overriding. It is this aspect of common-sense morality that I wish to provide a defence for.

In chapter one I set out my metanormative and metaethical position. This chapter also contains a preliminary and introductory discussion of some the over-arching themes in current thinking about practical reason and morality. I have also, for the sake of completeness, included a discussion of Prichard's intuitionism about morality. Chapter two, in a way, forms the bulk of this thesis. In it I outline, discuss, and eventually reject, Bernard Williams' internalism about practical reason. This chapter makes almost no use of the words 'morality' or 'obligation'. Rather, I focus on matters internal, so to speak, to internalism. In chapter three I carry on where chapter two left off: with more criticism of internalism. Here I focus on the connection between internal reasons for action and theories of motivation. I eventually conclude

that we should be cognitivists about motivation. So why this emphasis on internalism? The answer is that I have always found internalism and its intellectual forerunner, David Hume, to present the best, most plausible, case against not only the Kantian nuances of common-sense morality, but also against the currently most popular view of practical reason(s): externalism. Chapter four begins by posing a problem for externalists who think they have an answer to 'why should I be moral?' This puzzle takes the form of a tricky dilemma, and my discussion of this dilemma centres around the writings of one of the most influential externalists today - T. M. Scanlon. I reject Scanlon's answer to this dilemma along with his particular version of contractualism.

Although my own account has close affinities with Scanlon's it is clearly distinguishable from it. I argue that what might be called a buck-passing account of wrongness, fully worked out, has many advantages: it promises to explain, in a metaphysically and epistemologically acceptable way, why we should be moral. It also provides a clear alternative to consequentialism - which I see as counting in its favour - and it squares well with our moral intuitions. Although I acknowledge that more needs to be said in order to fully vindicate this account, the two aims of this dissertation is to attempt to clear the way for such an account, and to provide at least a rough sketch of what a plausible buck-passing account of wrongness should look like.

---

1Jackson, p. 118

## **CHAPTER ONE: A RATIONAL FOUNDATION FOR MORALITY?**

Two things fill the mind with ever new and  
increasing awe, the more often and steadily we  
reflect upon them: *the starry heavens above*  
*me and the moral law within me.*

- Immanuel Kant<sup>1</sup>

Once you give up integrity, the rest is a piece  
of cake.

- J. R. Ewing<sup>2</sup>

### **1.1 Introduction**

Why should I be moral? This question is probably as old as ethics itself. So too, it seems, is the worry that no satisfactory answer can be given to it. The question (albeit in slightly different versions) was put forward by several of Socrates' interlocutors. Callicles, Thrasymachus, and Glaucon all demanded a showing that it is in one's self interest to be moral (or just). The question is perhaps best, or at least most provocatively, formulated by Glaucon who

wants to know whether someone who is in the possession of the Ring of Gyges (and who as a result would be guaranteed impunity) still has good reason not to act immorally. Socrates' famous answer is that being an unjust person (or acting unjustly) comes at the price of having a fragmented or chaotic soul. Although an unjust person may be able to secure many goods for himself, having a well-ordered soul is the highest good a man can possess. It is interesting, and significant, that Socrates' answer has traditionally been interpreted as being unashamedly egoistic: it is in your self-interest to be just because it is in your self-interest to have a well-ordered soul - you would be worse off without one.

A considerable part of Western ethics after Plato has consisted in defending or rejecting Socrates' answer to what we might call the moral question ('why should I be moral?'). Very roughly, the sceptic wants to know why he should behave in the ways we usually call moral. And typically (and apparently), the only *kind* of answer that has seemed to be able to give the sceptic what he wants is an answer to the effect that it is in his self-interest to do so. Although Glaucon's question (and Socrates' answer) has worried moral philosophers for the last two thousand years, it is worth noticing that the same kind of worry or anxiety has never been aroused by the question 'Why should I do what is in my self-interest?'.

Many philosophers have taken it for granted that it is true that we ought to be prudent. Even those who deny this are

nonetheless happy to endorse the weaker claim that it is true that we have at least some reason (or a reason) to be prudent. This appears to have been the case for the majority of thinkers in Ancient Greece. But is it true? Perhaps. But without *argument* to support the supposed truth of 'you ought to be prudent', this claim seems no less problematic than the question we are concerned with. As Stephen Darwall points out: '...many of the same philosophical motives that prompt the question "why be moral?" should lead us to be similarly questioning of any other system of considerations and norms.'<sup>3</sup> In other words, even if we can show that prudence requires us to be moral, why is it so obvious that we ought to be prudent? Have we gained anything, philosophically speaking, by showing that considerations of prudence demand that we act morally (if this is indeed the case)? Prudence, as a system of norms, does not, on the face of it, have a more privileged normative status than morality. And, it seems, neither does any other system of norms.

Much of moral philosophy, then, has been devoted to establishing that morality can be given a *rational* foundation. In fact, many philosophers (past and present) think this should be moral philosophy's primary task. I think that in one sense of the word 'rational' this is true. However, and as I hope to show, this choice of terminology has not exactly been helpful in the search for an answer to the challenges of the sceptic. Given the ambiguity

(and possible vagueness) of the word 'rational' I think that in many ways the search for a rational foundation for morality has taken philosophy into places it need not go in order to provide an answer to 'why should I be moral?'. I hope that by the end of this chapter it will become clear what I mean by this.

The traditional worry, then, is that without a rational foundation, morality must be a chimera, a fantasy propagated by various moralists and religious folk throughout history. If there is no rational foundation for morality are we not better off without it? One author who thinks otherwise is John Mackie. He seems to suggest that although morality is indeed an illusion, we will all be better off if we act as if it were not.<sup>4</sup> Bernard Williams similarly claims that although 'the morality system' does provide us with moral obligations, by engaging in sound *ethical* reasoning we can come to see that these obligations are not as important as the morality system claims that they are.<sup>5</sup> Williams too writes that we would be better off without morality in its distinctively modern (and, as he sees it, Kantian) form; he advocates, if not a return to, then at least a favourable re-appraisal of, the ethical ideas offered by the ancient Greeks.

Nonetheless, for those of us who believe that there is a rational foundation to be found (not to say 'discovered'), one of the biggest obstacles to establishing that there is such a foundation is that there is widespread disagreement about what such a



foundation would be like - what it would consist in. And one of the reasons for *this* disagreement is that there is an equally widespread disagreement about what it means to say of a person, an action, or an institution that he or it is rational. So one of the aims of this chapter is to provide an outline of what, in the requisite sense, a rational foundation of morality would consist in. However, there is a more immediate concern we need to attend to first. If there is to be a rational foundation for *morality*, we need to be clear about what it means to be moral. Without a clear understanding of what it means to be moral we run the risk of missing our target; we could end up with an account of the rational foundation for something that looks very much like morality but which on closer inspection turns out to be mere schmorality.

## **1.2 'Being moral' and the choice of methodology**

It is no easy task to say what sort of a person a moral person is. I take it that a moral person is someone who has taken up 'the moral point of view'; someone who accepts that moral demands are what Kant called categorical imperatives (which is not meant to beg any questions against non-Kantian theories; I take it that utilitarians for example would agree that 'you ought to maximise happiness' is a categorical imperative in this sense). A moral person is someone who, among other things, accepts that moral demands are objective

and inescapable. Even more importantly, a moral person is someone who not only accepts these claims but who also structures her life in accordance with these moral demands; i.e. she is a person who accepts that if an act is wrong then it ought not be performed, and who, at least under normal circumstances, acts in accordance with this acceptance; i.e. she refrains from doing (or refrains from intending to do) what she believes to be wrong.

This sort of person should be contrasted with the kind of person who, in Kant's words, acts *merely in accordance* with morality, but not *for the sake* of morality. An example of this kind of person is Kant's shop keeper who decides not to overcharge his more gullible customers in order to maintain profits in the long run. Similarly, Hume's 'sensible knave' who believes '*honesty is the best policy*', may be a good general rule, but is liable to too many exceptions; and he...conducts himself with most wisdom, who observes the general rule, and takes advantage of all the exceptions'<sup>6</sup> is not someone we would normally call moral. Likewise, a person who, inspired by David Gauthier's *Morals by Agreement*<sup>7</sup> perhaps, manages to make himself 'translucent' and who succeeds in disposing himself to be such that he acts in accordance with common sense morality (provided that others are similarly motivated) can never be anything more than a sophisticated egoist. Since it is true of 'economic man' – no matter how sophisticated and translucent he may be – that were he able to

get away with lying, stealing, and cheating, he would do so were these activities to be in his self-interest. Such a person could not possibly be called moral, even though he would act in accordance with common sense morality in the vast majority of nearby possible worlds.

Now one of the difficulties in specifying what is involved in being moral consists in the fact that there is widespread disagreement about what morality actually demands that we do; in other words, there is widespread disagreement about which actions are right and which ones are wrong. This problem belongs to what is normally referred to as 'normative ethics' in introductory books on ethics, and I am not interested in trying to answer questions of this sort; or not directly interested in doing so anyway. I am interested in the question: 'Suppose we had a complete and correct list of actions that we are morally obliged to do and to refrain from doing. Why should anyone pay any attention to what is on that list?' As James Dreier puts it:

...suppose a moral theorist has proposed a certain set of rules, as the set of moral rules, the set that constitutes morality. We are bound to have some questions. We might question whether these rules really do constitute morality. This challenge is a kind of request for justification: what is the justification for the claim that

these rules constitute morality? But we could also ask for another kind of justification. We could ask what reason we, or anyone, has to follow these rules. This would be a request for a practical justification.<sup>8</sup>

So the idea is that we can distinguish two sorts of questions:

- (a) What is the morally right thing to do (in this situation)?
- and
- (b) Why should I care about the answer to the previous question?

(a) is a question about what we might call the content of morality and (b) is a question about the normative force of moral considerations. I understand the question 'why should I be moral?' as being equivalent to (b); it as a request for a practical justification of morality. This request would be satisfied by a showing that moral demands really are objective and normatively inescapable. But this of course prompts the question 'how can we know whether a certain set of demands are objective and inescapable unless we know *what those demands are?*'

Some philosophers believe that we cannot answer the latter question without having first answered the former question. According to these philosophers, questions about the normative force of various considerations characterised in general terms (such

as moral or prudential considerations) cannot be separated from the substantive content of those considerations. In opposition to this line of thought, some philosophers (usually of a Kantian persuasion) believe that we can only hope to derive, and ultimately justify, the content of morality by first asking questions of the second kind. So there are some methodological issues at stake here. Scanlon has recently commented on this choice of methodology. He says:

Explanations of the importance of morality and its reason-giving force can...be compared...according to their degree of formality or, on the other hand, of substantive content. The strategy of formal explanation is to appeal to considerations that are as far as possible independent of the appeal of any particular ends. Kant's theory is a leading example insofar as he undertakes to show that anyone who regards him- or herself as a rational agent is committed to recognizing the authority of the Categorical Imperative. [...] The alternative strategy is to explain the reason-giving force of moral judgements by characterizing more fully, in substantive terms, the particular form of value that we respond to in acting rightly and violate by doing what is wrong.

And, he goes on to say,

Formal accounts have been attractive because it has seemed that the force and inescapability of the moral "must" would be well explained by showing that moral requirements are also requirements of rationality, and not dependent on the appeal of any particular good. But although showing this might provide the secure basis that some have sought for the demand that everyone must care about morality, it does not give a very satisfactory description of what is wrong with a person who fails to do so. The special force of moral requirements seems quite different from, say, principles of logic, even if both are, in some sense, "inescapable." And the fault involved in failing to be moved by moral requirements does not seem to be a form of incoherence.<sup>9</sup>

What Scanlon calls the formal approach better suits my ambitions in this dissertation, but I disagree with his last point - I do think that the fault involved in failing to be moved by moral requirements *is* a form of incoherence. I shall of course have to make good on this claim in due course.

I think we can deflect some of the worries associated with the idea that we cannot answer questions about the normative force of moral considerations without knowing what the content of those

considerations are by focusing on the fact that although there is, undoubtedly, some disagreement about what morality demands, there is also considerable agreement on what morality *could not* demand on pain of ceasing to be 'a morality'. A 'morality' which allowed or required(!?) the torturing of infants for the sake of sadistic pleasure would not be considered a misguided or mistaken morality; rather it would not be considered a morality at all. If someone were to tell us that he believed that it was morally permissible for him to torture babies for fun, we would (I dare say) be warranted in believing that this person is either irrational, or that he is insincere, or that he does not understand the meaning of the locution 'morally permissible'. If morality demands anything at all, it certainly demands that we don't torture babies for fun. As Michael Smith puts it:

...there are platitudes concerning the *substance* of morality: 'Right acts are often concerned to promote or sustain or contribute in some way to human flourishing'... 'Right acts are in some way expressive of equal concern and respect'... What these platitudes about substance force us to admit, at the very least, is that there are limits on the kind of content a set of requirements can have if they are to be moral requirements at all, as opposed to requirements of some other non-moral kind....<sup>10</sup>



I think there is wide convergence in opinion on this matter. I also think this is so for a fairly large number of substantive moral claims. I am simply going to assume that the content of morality (whose overriding normative force I am interested in explaining) is relevantly similar to what we might call 'common sense' morality; where common sense morality consists in a set of moral beliefs that are widely shared and in whose truth we have a high degree of confidence.

### **1.3 Metaethics and metanormativity**

Although I am interested in (b) rather than (a), (b) in a way presupposes that there are answers to be given in response to (a). But why should we accept that there are such answers to be found? After all, there is a long standing tradition in ethics which denies that there are any answers to questions like 'what is the morally right thing to do?'. Generalising somewhat, we can say that antirealists (or irrealists) of various kinds (e.g. nihilists and error theorists) have all argued that since there are no moral properties in the world, there cannot be such a thing as 'doing that which is morally required'. Antirealists aren't the only sceptics however. As we shall see, H. A. Prichard believes that whereas (a) is a perfectly respectable question (to which a positive answer can be given), (b) is at best confused and at worst illegitimate. I shall discuss



Prichard's scepticism presently. First, however, I need to say something about the challenge presented by the antirealists.

I should begin by pointing out that the kind of scepticism (if I am allowed to call it that) represented by antirealists of various kinds in metaethics, can also be brought to bear on normative questions and discourse in general. If, like Mackie, we think that moral properties are too queer to be allowed into our ontology, why should we not think the same about normative properties in general? The reason why moral properties are queer, according to Mackie, is that they (supposedly) have this 'to-be-doneness' built into them. But if this is the ground for rejecting the existence of moral properties, is this then not also sufficient ground for rejecting the existence of normative properties (or relations) in general? After all, if the existence of moral properties is to be rejected because of the supposed queerness of their built-in 'to-be-doneness', then since normative properties or relations could quite plausibly be seen as *simply being* this 'to-be-doneness' relation, these would have to be rejected on the same grounds. Presumably, for the sake of consistency, if normative properties have to be rejected, they have to be rejected across the board. This would mean that the properties of, say, 'being such that it ought to be believed' and 'being a reason to believe' would have to be rejected as well. If this is so, however, none of Mackie's arguments can be said to be *rationaly* persuasive. Even if his arguments were sound,

we would have no reason whatsoever to assent to them. But if this is so, one might then (for good reason as it were) begin to wonder whether this is a coherent idea at all.

In light of this challenge, however, I need to say something about the assumptions I am going to make in this dissertation. To begin with, I am going to assume cognitivism about moral and normative discourse; i.e. I am assuming that moral and normative judgements are expressions of *beliefs* and are therefore 'truth-apt'. According to the 'minimal' theory of truth-aptness, a sentence is truth-apt if the following two criteria are met:

(a) *Discipline*: The sentences figure in an area of discourse that is disciplined. There must be standards operative with respect to which uses of those sentences are judged to be appropriate or inappropriate. There must be acknowledged standards for the proper and improper use of those sentences.

(b) *Syntax*: The sentences must possess the right sort of syntactic features. In particular, they must be capable of conditionalization, negation, embedding in propositional attitudes and so on.<sup>11</sup>

The idea is that if sentences in a given domain of discourse conform to these standards, then these sentences (and, presumably, the domain of discourse as a whole) are truth-apt; i.e. they are assessable in terms of truth and falsity. This should be contrasted with non-cognitivism (and those theories which fall under that umbrella concept, such as expressivism and emotivism), which holds that moral and normative discourse is not truth-apt; that moral and normative discourse or judgements should be taken to express or report agents' feelings or sentiments.<sup>12</sup>

I think it is quite clear that moral and normative language conform to the standards of minimalism. Sentences like 'It is wrong to tell lies' and 'Smith ought to look after his health' certainly meet the syntactical criterion. In fact, one of the most celebrated arguments against non-cognitivism – the so-called 'Frege-Geach Argument' – derives much, if not all, of its strength from the fact that moral judgements are capable of conditionalisation – a fact which non-cognitivists are still struggling to explain (away?).<sup>13</sup>

What about discipline? Again, it seems moral and normative discourse satisfy this criterion as well. Here I shall only point to few examples. It is a standard of both domains of discourse that claims made within these discourses are 'universalisable': if it is wrong for Bob to  $\phi$  in circumstance C, then it is wrong for anyone relevantly similar to Bob to  $\phi$  in circumstance C. Likewise in the (more general) normative domain: if Sarah ought, or has a reason to  $\phi$  in

circumstance C, then anyone who is relevantly similar to Sarah ought, or has a reason, to  $\phi$  in circumstance C. To claim otherwise would be to apply improper standards of use for moral and normative sentences. Other examples of such standards might include the prohibition against deriving 'ought' statements from 'is' statements and a licence to validly infer 'can' from 'ought' etc. Likewise, all deontic logics (as far as I am aware), as well as natural language, allow us to infer 'ought' from 'not permissible to not' and vice versa; it also allows us to infer 'ought not' from 'not permissible' and vice versa.<sup>14</sup>

Furthermore, and importantly (and as we saw in the quote from Smith) moral discourse is also disciplined by various platitudes. Although some of these have already been mentioned, Peter Railton lists the following platitudes<sup>15</sup>: moral discourse is surface cognitive – i.e. on its face it bears all the marks of cognitive discourse; statements of moral principles, and the stances taken in disagreements over such principles, are on their face nonrelativistic, or surface universal; moral judgements are typically linked to motivation in various ways – e.g. a judgement that X is wrong typically carries with it a motivation not to do X; moral judgements are surface categorical – i.e. ordinary lack of motivation or willpower on the part of the person to whom we are attributing a moral obligation is not an excusing condition but rather a fault, liable to sanction; moral discourse is surface justified by reasons of

well-being, nonpartiality, and cooperation – i.e. these are typically the kinds of considerations that ground moral judgements. This list is by no means meant to be exclusive or exhaustive. There may be (and probably are) many other moral platitudes besides the ones just mentioned. Even so, how do these platitudes help shape the discipline of moral discourse? I think the answer lies in the fact that all the platitudes listed above are prefaced by the word 'surface'. Moral discourse is disciplined by these platitudes in the sense that the plausibility of a moral theory is proportional to the number of platitudes it can accommodate satisfactorily. The failure of a theory to accommodate a particular platitude counts against the plausibility of the theory, unless of course defenders of the theory can persuasively show why we should simply discard the platitude in question. I think the same applies to theories of more generally normative concepts as well.

However, one of the main challenges for those who want to maintain that moral and normative discourses are truth-apt is to provide a plausible epistemology for moral and normative propositions. The non-cognitivist can of course supply such an epistemology but he can do so only at the cost of robbing normative and moral discourse of their truth-aptness. Since I am not going to argue for cognitivism, I shall not spend much time addressing the epistemological issue. Nonetheless, what I will say in chapter four will, I think, go a long way toward providing such an epistemology.

Unfortunately, most, if not all of this is compatible with what the error theorist wants to claim. The error theorist maintains that moral and normative discourse is truth-apt all right, but, he claims, since these discourses presuppose that there exists some mind-independent *sui generis* normative and/or moral sector of reality, all claims made within normative and moral discourse are false, since there simply is no mind-independent normative or moral order which is capable of conferring truth on these claims. My reply to the error theorist is that (insofar as this challenge can be made sense of) we can accept irrealism and cognitivism without thereby being committed to error theory. The idea behind the position I have in mind is that although judgements about moral and normative matters are expressions of beliefs (and which are, as a result, capable of being true or false), this does not imply that judgements about the moral or the normative are judgements about some section of reality which is independent of our cognising about it. As Scanlon puts it (in reference to discourse about reasons):

...in order for judgements about reasons to be taken to be about some subject matter independent of us in the sense required for it to be possible for us to be mistaken about them, what is necessary is for there to be standards for arriving at conclusions about reasons. ... It is not

necessary, in order to explain the possibility of being mistaken to construe the relevant subject matter in a metaphysical way as existing outside us. [...] The question of whether there are standards of the required sort is a substantive one within the subject in question...it need not be a metaphysical question about what exists....<sup>16</sup>

The distinctive error theoretical claim, as I see it, is that the truth of irrealism about ethics *makes it the case that* all ethical judgements are false. I suspect that the reason why error theorists believe this is that they (implicitly perhaps) subscribe to a correspondence theory of truth. If this theory is rejected, however, it is perfectly possible to remain an irrealist cognitivist without being an error theorist.

I cannot defend the claim that we should reject the correspondence theory of truth (at least with respect to the moral and normative domains) in favour of a minimal theory of truth(aptness). Here I am simply suggesting that by rejecting the correspondence theory of truth we can perfectly well remain irrealists about normativity and morality (i.e. we can deny that there exists some mind-independent, truth-making order of normative or moral facts) and we can remain cognitivists about normativity and morality (i.e. we can assert that normative and



moral discourse is truth-apt) without thereby committing ourselves to some form of error theory. I am simply going to assume that some moral and normative propositions are true and that these true propositions are not merely negative existential claims of the form 'It is true that there are no moral properties' or 'It is true that there are no reasons for doing anything'. Rather, I take it that there are true (atomic) propositions of the form 'It is wrong to steal' and 'Bob ought to look after his health'. The position I am assuming may be called 'non-error theoretic irrealist cognitivism'.<sup>17</sup>

It is against the backdrop of these admittedly large assumptions that I want to try to answer the question 'why should I be moral?' Since I cannot defend these assumptions, I have to admit that the conclusions I reach will be no stronger or more persuasive than the assumptions on which these conclusions stand. With these preliminaries in place, and before I outline what I think a rational foundation for morality (in the requisite sense) must consist in, I first need to say something about Prichard's view that 'why should I be moral?' is an illegitimate question.

#### **1.4 Prichard's intuitionist challenge (1)**

In his essay 'Does Moral Philosophy Rest on a Mistake'<sup>18</sup> H. A. Prichard claims that, traditionally, moral philosophy has consisted



more or less in an attempt to provide an answer to the following question:

Is there really a reason why I should act in the ways in which hitherto I have thought I ought to act? May I not have been all the time under an illusion in so thinking? Should I not be justified in simply trying to have a good time?<sup>19</sup>

Since he believes this is an illegitimate question, Prichard believes that moral philosophy rests on a mistake:

I...contend that the existence of the whole subject, as usually understood, rests on a mistake, and on a mistake parallel to that on which rests, I think, the subject usually called the theory of knowledge.

In this and the next section I will try to show that it is Prichard who is mistaken. The two questions asked in ethics and epistemology respectively, are legitimate questions. Prichard presents two arguments that supposedly show why the question above is illegitimate. One proceeds via an analogy with epistemology (and the mistakes supposedly made therein), and the other tries to show, more directly – within the confines of moral philosophy, so to

speak – why the question is illegitimate. I shall begin by discussing his analogy between moral philosophy and epistemology.

Prichard says:

Just as the recognition that the doing of our duty often vitally interferes with the satisfaction of our inclinations leads us to wonder whether we really ought to do what we usually call our duty, so the recognition that we and others are liable to mistakes in knowledge generally leads us, as it did Descartes, to wonder whether hitherto we may not have been always mistaken. And just as we try to find a proof, based on the general consideration of action and of human life, that we ought to act in the ways usually called moral, so we, like Descartes, propose by a process of reflection on our thinking to find a test of knowledge, i.e. a principle by applying which we can show that a certain condition of mind was really knowledge, a condition which *ex hypothesi* existed independently of the process of reflection.<sup>20</sup>

In the epistemological case, then, when we attempt to provide a proof of the alleged fact that a given mental state was one of knowledge we set ourselves the task of finding a criterion by which we can judge whether we are in possession of genuine knowledge.

And it is only by knowing that a given mental state satisfies this criterion of knowledge that we can be said to have genuine knowledge:

instead of its being the fact that the knowledge that *A* is *B* is obtained directly by consideration of the nature of *A* and *B*, the knowledge that *A* is *B*, in the full or complete sense, can only be obtained by first knowing that *A* is *B*, and then knowing that we knew it by applying a criterion, such as Descartes principle that what we clearly and distinctly conceive is true.<sup>21</sup>

But, Prichard tells us, this epistemological enterprise is mistaken:

Now it is easy to show that the doubt whether *A* is *B*, based on this speculative or general ground, could, if genuine, never be set at rest. For if, in order to know that *A* is *B*, we must first know that we knew it, then really, to know that we knew it, we must first know that we knew that we knew it.<sup>22</sup>

So an infinite regress threatens this kind of argument, and this is one of the fundamental mistakes Prichard wants to attribute to Descartes and, consequently, to the epistemology as a whole.

Whether Descartes actually made this mistake or not is beyond the scope of the present discussion. Prichard nonetheless seems to want to draw the conclusion that since we cannot (on pain of regress) include a criterion of this sort (or indeed any other) in our definition of knowledge, knowledge must be *sui generis*; like goodness is for Moore<sup>23</sup>, knowledge, for Prichard, is unanalysable.

The epistemological mistake, Prichard maintains, is the result of our inattention to the 'immediacy of knowledge':

[A moral] apprehension is immediate, in precisely the same sense in which a mathematical apprehension is immediate, e.g. the apprehension that this three-sided figure, in virtue of its being three-sided, must have three angles. Both apprehensions are immediate in the sense that in both, insight into the nature of the subject directly leads us to recognise its possession of the predicate; and it is only stating this fact from the other side to say that in both cases the fact apprehended is self-evident.<sup>24</sup>

A striking feature of this passage is that the two kinds of knowledge it mentions are species of *a priori* knowledge. Mathematical judgements are paradigmatically *a priori*, and so too it seems, are moral judgements (although some philosophers working in the empiricist or naturalist tradition might think otherwise). However,

not all *a priori* judgements are immediate in this sense. A difficult proof in logic for instance sometimes requires careful examination before a correct verdict regarding its validity can be reached. Hence, there is a significant difference between *a priori* knowledge and immediate knowledge. So the 'immediacy' of knowledge is something we should be sceptical about, and this (justified) scepticism, I think, should be extended to the claim that knowledge and certainty are inextricably linked as well.

It is a mistake to think that we have to be either Cartesian internalists or intuitionists about knowledge. There is another interpretation of the Cartesian programme; an interpretation which escapes the infinite regress objection. Consider Descartes' principle:

- (A) For all *P*, if I clearly and distinctly perceive that *P*, then I know that *P*.

Contrary to Prichard's interpretation of Descartes, (A) could be understood as not being a *ground* (or criterion) for knowledge, rather, it can be regarded as a *source* of knowledge. As Van Cleve points out:

In order to [know] a proposition I do not need to know that I am clearly and distinctly perceiving it, nor that

whatever I so perceive is either certain or true. It is enough that I *do* clearly and distinctly perceive the proposition. (A) says that this is enough. . . . (A) is not a principle I have to *apply* in order to gain knowledge, I need only *fall under* it.<sup>25</sup>

Thus there is a 'criterion-based' theory of knowledge which does not obviously lead to an infinite regress. This has become known as 'externalism' in epistemology. Within the class of externalist theories of knowledge there are several proposals available as to what this criterion is. Alvin Goldman, for instance, has advocated a theory of knowledge according to which the criterion under which 'one has to fall' is that one's belief that p was *caused* by p.<sup>26</sup> Similarly, Robert Nozick has constructed a conditional theory of knowledge according to which one's belief that p must be such that, counterfactually, had p not been the case, one would not have believed p, and were it the case that p, one would believe that p.<sup>27</sup> Both Goldman's and Nozick's theories are sometimes called 'reliabilist' theories as they both, in their respective ways, attempt to specify ways in which we gain knowledge through a 'reliable knowledge-generating process'.

The crucial difference between these theories and the ones Prichard has in mind, is that, according to externalists, it is not a necessary condition for S's knowing that p that S knows that he

knows that *p*. It is sufficient for *S* to know that *p* that *S* believes *p*, that *p* is true, and that *S* has come to believe *p* through some reliable process. *S*'s having acquired the true belief that *p* in the right way does not, however, mean that *S* is *justified* in believing that *p* or asserting that *p*. But this, in turn, can be taken to show that justification is not essential to knowledge.

Although I will not try to defend externalism here, it should be pointed out that it has some notable advantages over internalism. Perhaps the most appealing aspect of externalism is that it can accommodate the common-sense conviction that higher animals, young children, and unsophisticated adults can possess knowledge without necessarily knowing that, or why, they do. Although this sort of theory might not, in the end, save Descartes from the fate Prichard assigns him, it nonetheless provides us with a *prima facie* alternative to both Descartes and Prichard.

So Prichard cannot simply help himself to the conclusion that if Cartesian internalism fails intuitionism is the only available alternative. By divorcing the notions of justification and knowledge we can maintain that there is a third position available, a position that Prichard has not considered. So Prichard faces a dilemma here. Either he has to retract his claim that the epistemological case and the moral case are analogous, thus eliminating one of his main arguments in support of the claim that the question asked by moral philosophy is illegitimate, or he has to admit that a non-

intuitionistic answer to 'Is there really a reason why I should...?' is still on the table.

### **1.5 Prichard's intuitionist challenge (2)**

Consider again Prichard's question:

Is there really a reason why I should act in the ways in which hitherto I have thought I ought to act? May I not have been all the time under an illusion in so thinking? Should I not be justified in simply trying to have a good time?

I suspect that Prichard thinks that this question is the same as the question 'why should I be moral?'. This is not so however – they are different questions. For one thing, the question Prichard has identified is in fact not one, single question, but three separate questions. To distinguish 'why should I be moral' from Prichard's questions, I will refer to the former as 'the moral question' and the latter as 'Prichard's question' (or as 'his question'). If one of the chief aims of moral philosophy 'as usually understood' is to answer Prichard's question (understood as a conjunction) then Prichard's conclusion is absolutely correct - moral philosophy does rest on a mistake.



I believe 'why should I be moral?' is a legitimate and important question. However, Prichard's failure to establish the illegitimacy of his question is instructive in several ways; we can learn a great deal about how one should go about answering the moral question by pointing out where and why Prichard is mistaken. In the end, or so I will argue, it is Prichard's question and his subsequent arguments, not moral philosophy, that rest on, not only a mistake, but several mistakes.

As I said, Prichard's question is in fact three separate questions. They are:

- (I) Is there really a reason why I should act in the ways in which hitherto I have thought I ought to act?
- (II) May I not have been all the time under an illusion in so thinking?
- (III) Should I not be justified in simply trying to have a good time?

Intuitively, the answers to these questions are 'that depends on what you hitherto thought you ought to do', 'yes', and 'probably not'. At this stage it is far from obvious that the second and third questions have any bearing whatsoever on the first question. Let us set these consideration to the side and take a look at Prichard's 'reductio' answer to his question. He says:

[Answers to this question] all fall, and fall from the necessities of the case, into one of two species. *Either* they state that [1] we ought to do so and so, because, as we see when we fully apprehend the facts, doing so will be for our good, i.e. really, as I would rather say, for our advantage, or, better still, for our happiness; *or* they state that [2] we ought to do so and so, because something realized either in or by the action is good. In other words, the reason 'why' is stated in terms either of the agent's happiness or of the goodness of something involved in the action.<sup>28</sup>

For simplicity's sake I have labelled these alternatives [1] and [2], respectively. [1] attempts to supply each agent with an egoistic (agent-relative) reason for acting morally. [2] is more agent-neutral in spirit.<sup>29</sup>

We should note right away that there is an important assumption underlying this passage: all potentially acceptable answers to his question necessarily fall into one of the two categories described by Prichard. By this I take Prichard to mean that answers which fall under these categories are the *only* ones that could possibly supply the type of answer we are looking for. There are two obvious questions at this point: (1) are these types of

answers really the only ones available?; and (2) are these types of answers really incapable of answering his question? I think the answer to (1) is no. The two categories Prichard identifies are too narrowly *teleological* (or value-based). Since agent-relative and agent-neutral reasons exhaust the logical space of reasons (so described), it is true that any acceptable answer to Prichard's question must fall within one of *these* categories. This does *not* mean, however, that all reasons (agent-relative and agent-neutral) must be teleological reasons. The distinction between agent-neutral and agent-relative reasons is not the same distinction as that between teleological and non-teleological reasons. So Prichard's first mistake consists in his thinking that any acceptable answer to his question must proceed via a teleological justification of morality. Let us continue by looking at Prichard's answer to (2).

Prichard seems to suggest that his question should be read as a mistaken request for a proof of the truth that we are under moral obligations. According to Prichard, it is a question asked by someone who is no longer confident that he ought to do those things he previously thought he was under an obligation to do. It is fairly clear that the obligations Prichard has in mind here are the various obligations associated with acting in ways we normally describe as being in accordance with morality: keeping ones promises, repaying debts, telling the truth etc.

However, it is unclear (at this point anyway) what the content of this person's doubt is. Prichard does not explicitly tell us whether the person asking the question has come to doubt whether a) it *really is* morally obligatory to do these things; or, b) whether, although these acts may be morally obligatory, it is reasonable to do what morality tells him to do; i.e. whether moral obligations really make normative claims on us. So we need to find out which kind of doubt Prichard is concerned with in order to understand what it is he wants a proof of. The two doubts are after all very different. The following passage suggests that Prichard's question should be understood as an expression of a doubt of the latter kind:

The formulation of the question implies a state of unwillingness or indifference towards the action, and we are brought into a condition of willingness by the answer. And this process seems to be precisely what we desire when we ask, e.g., 'Why should we keep our engagements to our own loss?'; for it is just the fact that the keeping of our engagements runs counter to the satisfaction of our desires which produced the question.<sup>30</sup>

This passage suggest that the question is being asked by someone who wants a psychologically compelling answer. But this is only

one possible motive behind asking this question. We could also ask this question because we want to enquire about the normative status of moral obligations - is it really true that we ought to do what we are morally obliged to do? If desire satisfaction were the only source of normativity then the two questions would be asking for the same thing, but why should we accept that this is the case? The question may be asked, for example, by someone who has come to doubt whether moral obligations are genuinely normative, not because of his 'appreciation of the degree to which the execution of these obligations is contrary to our interest', but because he has become convinced that, say, aesthetic considerations bid him to act contrary to the demands of morality.

If we interpret (I) in such a way that it presupposes that any acceptable answer to it must be psychologically compelling, then since it is doubtful that we should be able to provide an affirmative answer to (I) that every person will find psychologically compelling, I think we can safely say, with Prichard, that (I) is illegitimate. But why should we interpret (I) in this way? As Parfit puts it, our aim here may be 'not influence, but truth.'<sup>31</sup>

Given Prichard's assumption that the only kind of reason(s) that can be given for why we ought to act in accordance with morality are necessarily derived from teleological considerations, then if the 'desire satisfaction' answer is unacceptable<sup>32</sup>, the only other alternative, according to Prichard, is to base 'the obligation to

do something on the *goodness* either of something to which the act leads or of the act itself ([2]).<sup>33</sup> The most common version of this alternative is of course utilitarianism 'in the generic sense' (to borrow Prichard's term). So let us assume (with Prichard) that utilitarianism supplies such an answer: those acts are right which produce maximum happiness. What we would like to know next is whose happiness we should be concerned with. Again, for the same reasons as mentioned above, if utilitarianism is to be able to account for the sense of being under an obligation that Prichard is after, it cannot, according to Prichard, appeal to 'our own happiness' - that is to the happiness of the agent who finds himself under an obligation to promote happiness. If it did the answer would resolve itself into the same answer as [1]. So what are the alternatives available to the utilitarian? According to Prichard, there are only two alternatives available: either

[2.1] Anyone's happiness is a thing good in itself and  
therefore we ought to do whatever will produce it; or,

[2.2] Working for happiness is itself good, and the intrinsic  
goodness of such an action is the reason why we ought  
to do it.<sup>34</sup>

These two options, then, are the only available ones left. What, if anything, can be said in favour of these kinds of positions? Prichard says:

The advantage of this appeal to the goodness of something consists in the fact that it avoids reference to desire, and instead refers to something impersonal and objective. In this way it seems possible to avoid the resolution of obligation into inclination.<sup>35</sup>

Prichard equates [2.1] with utilitarianism 'in the generic sense'. According to Prichard, utilitarianism 'takes its stand upon the distinction between something which is not itself an action but which can be produced by an action and the action which will produce it, and contends that if something which is not an action is good then we *ought* to undertake the action which will, directly or indirectly, produce it.'<sup>36</sup> But for this argument to succeed (i.e. to restore the sense of obligation), Prichard continues, we must assume not only that 'what is good ought to be', but also that the apprehension that a good state of affairs ought to be involves the feeling of imperativeness or obligation which is to be aroused by the thought of the action which will originate it - otherwise 'the argument will not lead us to feel the obligation to produce it by the action.'<sup>37</sup> But, Prichard argues, both these assumptions are false.

First, the word 'ought', he claims, refers to actions and actions alone.<sup>38</sup> He explains:

Even if we are sometimes moved to say that the world or something in it is not what it ought to be, what we really mean is that God or some human being has not made something what he ought to have made it.<sup>39</sup>

I think this is right, or roughly right anyway. 'Ought' statements do in a sense apply to actions, but they also apply to *agents*. More specifically, 'oughts' are relational, they hold (like reasons) between agents and actions (and circumstances perhaps). Of course we say things like 'the train ought to have arrived by now', but 'ought' is here used in a different sense than in a statement like 'you ought to be kind to your little brother'. In the first instance the word 'ought' is used in what we might call a predictive sense whereas in the second it is used in a normative sense. It is in this second sense that 'ought' only applies to agents. It does not make sense to say that the world ought to be such and such (in the normative sense). If it were true that the world ought to be such and such but that it is not, then the world would be a thing which is in some way criticisable (or blameworthy perhaps), and clearly the world is not. The world is not an agent; it cannot be held responsible for 'what it does'.



However, it is hard to see why a teleologist should be bothered by this. No doubt teleologists frequently disagree about what is intrinsically good; happiness, well-being, pleasure, and human perfection are only some suggestions. They also disagree about what the appropriate response to these suggested goods should be; some advocate maximisation while others endorse 'satisficing' accounts as the appropriate response to the value in question.<sup>40</sup> But this does not mean that teleologists are committed to the claim that 'what is good ought to be'. Rather, they believe that what is (maximally, satisficingly etc.) good ought to be brought about. Here the 'ought' *is* attributed to agents, and bringing about a particular state of affairs *is* (or at least can be) an action. Furthermore, teleologists need not concede that 'what is good ought to be brought about' is *derived* from 'what is good ought to be'. Prichard seems to assume that teleologists must conceive of the good/right relation in the following way<sup>41</sup>:

- (1) A certain state of affairs, *x*, is the best state of affairs.
- (2) Therefore I ought to bring about *x*.

We can agree with Prichard that (2) does not follow from (1). But this does not mean that the teleologists must presuppose 'What is best ought to be'. (Were they to do so, the argument would still be invalid.) They need not suppose anything of the sort. Rather, they

could (and presumably do) claim that what the argument above presupposes is 'I ought to bring about the best state of affairs'. With this assumption in place, (2) does follow from (1).

What about the second assumption - the assumption that 'the apprehension that a good state of affairs ought to be involves the feeling of imperativeness or obligation which is to be aroused by the thought of the action which will originate it'? Prichard's rejection of this assumption is based on his rejection of the first one: our feeling of imperativeness, or of being under an obligation, can never be aroused simply by the recognition that something is best. Since he believes that teleologists must hold that what is best ought to be, and that this ought does not refer to an action (but to a state of affairs), he says:

...it is merely stating another side of this fact to urge that we can only feel the imperativeness upon us of something which is in our power, for it is actions and actions alone which, directly at least, are in our power.<sup>42</sup>

This challenge has already been stopped dead in its tracks. If the teleologist simply rejects the assumption 'what is best ought to be' in favour of 'what is best ought to be brought about', Prichard's challenge loses all its force.

Nonetheless, I think it is fairly clear that Prichard means to say that teleology in general, and utilitarianism in particular, are, according to common-sense morality, counterintuitive. He says:

It is because this view is so plainly at variance with our moral consciousness that we are driven to adopt the other form of the view, viz. that the act is good in itself and that its intrinsic goodness is the reason why it ought to be done.<sup>143</sup>

Not surprisingly, Prichard dismisses the 'intrinsic goodness' answer as well. Our use of the term 'good', he says, is always in respect of the motive and refers to actions which have been actually done and of which we think we know the motive. 'Further', he says, 'the action of which we approve and which we should describe as intrinsically good are of two and only two kinds. They are either:

[2.2.1] Actions in which the agent did what he did because he thought he ought to do it; or,

[2.2.2] Actions of which the motive was a desire prompted by some [intrinsically] good emotion such as ... benevolence.<sup>144</sup>

Prichard does recognise that actions can be carried out as a result of a combination of these motives. Although such a combination of motives might present a problem with regards to the possible overdetermination of (operative, or explanatory) motives capable of producing an action, I think Prichard is right when he says that problems of this kind will not affect the argument he puts forth next. Before we take a look at this argument, let us remind ourselves about what Prichard has done so far.

Prichard began by considering the attempt to ground the reason why we ought to act in accordance with morality in agent-relative considerations (advantage, happiness etc.). 'But', he concludes, '[this] answer is, of course, not an answer, for it fails to convince us that we ought to keep our engagements; even if successful on its own lines, it only makes us *want* to keep them.'<sup>45</sup> This left us with a choice between grounding moral obligations in either the intrinsic goodness of universal happiness or in the intrinsic goodness of performing certain acts. The 'universal happiness' answer is dismissed because it is so 'plainly at variance with our moral consciousness'. So the only alternative left is to ground the reason why we ought to act in accordance with morality in the intrinsic goodness of *working* for happiness. But, Prichard says

...our approval and our use of the term 'good' is always in respect of the motive and refers to actions which have been actually done and of which we think we know the motive. Further, the actions of which we approve and which we should describe as intrinsically good are of two and only two kinds. They are either actions in which the agent did what he did because he thought he ought to do it, or actions of which the motive was a desire prompted by some good emotion....<sup>46</sup>

I think Prichard is right about this. I will have more to say about this issue later, but for an event to be an action that event must be under the conscious control of some agent - i.e. the event must be brought about intentionally, it must have some purpose. So when we evaluate the goodness of *actions* (as opposed to 'mere happenings' or 'quasi-actions'), we cannot accurately do so without knowing what the intentions were that brought about a particular event (a person's moving his body in a certain way). However, according to Prichard, this presents us with a dilemma:

The dilemma is this: If the motive in respect of which we think an action good is the sense of obligation, then so far from the sense that we ought to do it being derived from our apprehension of its goodness, our apprehension of its

goodness will presuppose the sense that we ought to do it. In other words, in this case the recognition that the act is good will plainly *presuppose* the recognition that the act is right, whereas the view under consideration is that the recognition of the goodness of the act *gives rise* to the recognition of its rightness. On the other hand, if the motive in respect of which we think an action good is some intrinsically good desire, such as the desire to help a friend, the recognition of the goodness of the act will equally fail to give rise to the sense of obligation to do it. For we cannot feel that we ought to do that the doing of which is *ex hypothesi* prompted solely by the desire to do it.<sup>47</sup>

There may be various ways for teleologists to respond to this challenge, but on the standard definition of teleology (i.e. that the goodness of an action can be defined independently of its rightness, and that rightness of an action is then defined in terms of some suitable response – maximising, satisficing etc. – to the good), Prichard has a real target here. Even so, this does not mean that Prichard has shown that the question he asked is illegitimate. At best, he has shown that teleological answers in general (and perhaps consequentialist answers in particular) will, at best, yield counterintuitive answers to his question, at worst they cannot

provide an answer at all. So much the worse for teleology perhaps. But, alas, there are other moral theories besides teleology.

In the end, I believe we must conclude that Prichard's arguments fail to establish that the moral question is illegitimate. His argument presupposes that there is a close analogy between the mistakes supposedly committed in both epistemology and in moral philosophy. But, as I have shown, no mistakes have been committed in either case. Prichard's attempt to establish the truth of intuitionism in moral philosophy by appealing to the supposed truth of intuitionism in epistemology does not succeed. The reasons for this is that Prichard has overlooked the fact that the forced choice between Cartesian internalism and intuitionism is a false dichotomy. Prichard is nonetheless right in believing that his question is illegitimate although he is not right about the reason why it is illegitimate. The illegitimacy stems not from the question as such, but rather from the presupposition that any satisfactory answer to it must be able to meet two particular criteria: it must be psychologically compelling, and it can only successfully proceed via an appeal to considerations of goodness. We need not accept either of these criteria.

## 1.6 Different conceptions of 'rational'

I said earlier that the question 'why should I be moral?' has traditionally been interpreted as a request for a showing that there is a rational foundation for morality. Much of the controversy surrounding the possibility of there being such a foundation for morality stems from the fact that the word 'rational' is ambiguous. In what follows I shall try to distinguish (some of) these senses of 'rational'.

Since there is an enormous amount of literature on various theories of rationality, I cannot possibly canvass all these theories in what follows. As a result I shall have to be selective. Theories of rationality can be divided into various categories depending on what one takes to be their most salient features, e.g. instrumental or non-instrumental theories, consequentialist or non-consequentialist theories, maximising or non-maximising theories etc., or into almost any combination of these. I shall, however, discuss a slightly different categorisation which cuts across many of the distinctions drawn above. I am going to call these formal and substantive theories of rationality, and I shall discuss them presently.

I should point out that the difference between the two theories I am about to discuss is unfortunately obscured by the fact that these theories are often referred to by other names in the literature. Scanlon, for instance, makes a distinction between a



'broad' and a 'narrow' conception of rationality, Parfit talks of 'substantive' and 'procedural' conceptions of rationality,<sup>48</sup> and Rawls distinguishes between 'reasonableness' and 'rationality' (a distinction, he claims, that traces back to Kant).<sup>49</sup> Although these distinctions do not exactly match each other, for our purposes we can say that Scanlon's broad conception of rationality, Parfit's procedural conception, and Rawls' sense of 'rational' are fairly close versions of the formal conception I have just been discussing.

Before we take a look at the substantive and formal conceptions of rationality, it will be useful first to consider the question 'what things can be said to be rational?' Well, quite a few it seems: actions, persons, dispositions, rules, norms, and institutions (to name but a few) all seem to be 'things' to which we sometimes attach the predicate 'is rational'. For instance, when someone asks us for advice about what to do we sometimes answer by saying that the rational thing *to do* is such and such. Similarly we often say of someone who acts in a certain way that she is a rational *person*. In both cases the predicate 'is rational' is invoked, but the kinds of objects this predicate is (or can be) true of are different.

Although we commonly attribute rationality to all sorts of things, I think that, on reflection, we should say that the primary 'bearers' of rationality are agents; more accurately we should say that agents are rational in so far as they possess, and to some

adequate degree, exercise certain capacities – namely the capacity to form normative judgements and the capacity to act (intentionally) in accordance with these judgements. Actions are in turn rational if they are caused by an agent's exercising these capacities. There are of course several kinds of normative judgements, and rationality and irrationality cannot simply be a matter of acting in accordance with any old normative judgement. Suppose Sue judges that she has a reason to  $\phi$ . Here it might be tempting to say that Sue would be acting rationally if she  $\phi$ -ed, but this is not necessarily so. Suppose Sue also judges that she ought not  $\phi$  (which, I take it, is a 'stronger' normative judgement than the judgement that she has a reason to  $\phi$ ). In this case, if she were to act in accordance with her judgement that she has a reason to  $\phi$  (i.e. if she  $\phi$ 's), then given that she also judges that she ought not  $\phi$ , she is not acting rationally.

So the required connection between her normative judgements and her actions (or, more accurately perhaps, her intentions) must be that she acts on her 'strongest' normative judgement. The word 'strongest' here is deliberately vague because I do not yet want to commit myself to any particular story about what the content of these judgements must be. I should however point out that 'strongest' here does not mean 'most firmly held' or something similar. Rather, an agent's strongest normative judgement should be understood as that judgement of the agent

whose content has the strongest normative modality: to act rationally Sue must act in accordance with her judgements about what she *ought* to do; about what she has *decisive* reason to do; about what she has *most* reason to do etc. Of course it might be the case that the contents of these judgements are the same. To judge that one ought to  $\phi$  might simply be the same thing as judging that one has sufficient reason to  $\phi$  etc. Since I will discuss these issues in some detail in chapter four, I shall simply leave the idea of acting in accordance with one's strongest normative belief intuitive at this stage. We may call this the 'formal' conception of rationality: an agent is rational if and only if she acts in accordance with her strongest normative judgement. The reason why the term 'formal' is apt here is that this conception of rationality says that rationality is, essentially, nothing over and above consistency. There are no constraints imposed on, or related to, the normative judgements themselves.

There are a few complications here however. First, a person may occasionally act irrationally without thereby deserving to be called an irrational person. I may have been weak willed on a few occasions (i.e. I may have failed to act in accordance with my strongest normative judgement), but this is not sufficient to make me an irrational person. Rational persons sometimes act irrationally. However, someone who acts irrationally on a sufficient number of occasions cannot, in any ordinary sense of the word, be

called a rational agent. Whether or not this means that 'rational' is a vague concept I shall leave unexamined as it is of no great importance in the present context.

Second for the formal conception of rationality to be plausible - indeed, for any conception of rationality to be plausible - we also need to add that the connection between an agent's normative judgments and her actions (or intentions) is not merely accidental. Consider the following case (due to John Broome). Suppose you believe (for whatever reason) that the hotel you're in is on fire. This belief (along with a desire to stay alive perhaps) might lead you to believe that you ought to jump into the canal outside the window. Suppose however that '[y]our belief that you ought to jump might perversely cause you to sing happily, so that you start to skip about on the windowsill and by accident jump into the canal.'<sup>50</sup> In this case your action is consistent with your strongest normative belief (that you ought to jump), but this would not be an example of a rational action. In order for an act to be rational, the action must come about as a result of, or because of, the (strongest) normative belief.<sup>51</sup> In other words, there must be some non-accidental connection between the agent's strongest normative belief and his subsequent behaviour. I shall take this 'non-accidental' rider as read in what follows.

Third, although I have said that the formal conception of rationality demands that there be a non-accidental connection

between an agent's strongest normative judgement and her action(s), this is, strictly speaking, not true. Suppose Anne believes that she ought to study for her exam but she doesn't do it. The fact that she failed to study for exam cannot by itself be enough to warrant a charge of irrationality. Perhaps she was prevented from studying by some 'external' force like an earthquake.

Considerations like these should force us, I think, to say that a person is rational only if she *forms the intention* to do that which she believes she ought to do.<sup>52</sup> However, since we are so used to talking about rational actions as if they are overt bodily motions, I shall continue to do so, even if it is, strictly speaking, not true.

The formal conception of rationality is, however, open to a fairly common objection. The objection to the formal account of rationality (and the similar ones just mentioned) is brought out by the following quote from Scanlon:

Beyond [acting in accordance with one's strongest normative judgement], are there also more substantive standards that it is irrational to violate? Is it sometimes irrational to fail to accept certain considerations as reasons? It seems to me that philosophical usage, but perhaps not "ordinary" usage, is divided on this point.<sup>53</sup>

This brings us to what we might call the 'substantive' conception of rationality. According to this conception of rationality, an agent's acting in accordance with his or her strongest normative belief(s) is a necessary but not a sufficient condition for an agent's being rational. On the formal conception, Sue's  $\phi$ -ing constitutes a rational act if and only if her  $\phi$ -ing is consistent with her strongest normative belief *regardless of what the content (or degree of warrant) of that normative belief is*. This is the feature of the formal conception of rationality that those who favour a substantive conception object to. To illustrate the difference between the two conceptions, consider a version of a famous example from Hume: The world is about to be blown up and Anne can prevent this from happening (and Anne knows that she can do this). Anne, however, believes she has most reason to scratch her little finger and that she has only a very weak reason to prevent the world from being blow up. On the basis of these beliefs, she stands there watching the world being destroyed while scratching her little finger. Is Anne acting rationally?

According to the formal conception of rationality Anne is acting rationally; her actions are consistent with her strongest normative beliefs (and, we may suppose, her actions are properly connected with those beliefs). If we should criticise Anne for anything, on the formal conception of rationality, the appropriate criticism is that she is (seriously!) *substantively mistaken* about

what she ought to be doing: how could it possibly be *true* that she has more reason to allow the destruction of the world than she has reason to refrain from scratching her finger?

According to the substantive conception, however, anyone who is as grossly substantively mistaken as Anne cannot avoid a charge of irrationality by simply pointing out that her action is consistent with her strongest normative judgement. Anyone in their right mind should recognise that there is more reason to prevent the destruction of the world than there is to scratch one's little finger. A failure to see this is plainly irrational. So even if Anne is acting consistently with her strongest normative judgement, the content of her judgement is so incredible that it warrants a charge of irrationality: believing that one has more reason to scratch one's little finger than to prevent the planet from being destroyed is simply crazy.

So what exactly is the substantive conception of rationality? This is a very hard question to answer since there is considerable disagreement (even among those who reject the formal conception of rationality) about what further conditions, beyond consistency, have to be satisfied in order for an agent to act rationally. For our purposes, and for reasons which will become clear presently, we need not answer this question. We can say that on a weak version of the substantive account, an agent's strongest normative judgement has to be warranted to some degree in order for her



action to be rational. On a stronger version of the theory, the agent's strongest normative judgement has to be *true* (and, plausibly, warranted) in order for her action to be rational.

Of the three accounts of rationality, or of rational action, I have just (briefly) discussed, the weak version of the substantive conception of rationality seems the most plausible to me. It simply sounds implausible that in order to act rationally, our strongest normative judgement has to be true. If I have a high degree of warrant for my belief that I ought to  $\phi$  and I  $\phi$  as a result of this, surely this is sufficient to make my action rational! To insist that I cannot act rationally unless my strongest normative judgement is true is simply too demanding; this is not what we mean by 'rationality' or 'rational action' in everyday talk. On the other hand, merely acting consistently with one's strongest normative belief does not appear to be sufficient for rational action either. However, I think the strong version of the substantive conception of rationality gives us exactly the sense of 'rational' philosophers have had in mind when they have been trying to secure a rational foundation for morality.

If we mean by 'rational' something similar to that denoted by the formal conception of rationality, we should not expect there to be a rational foundation for morality in the requisite sense. If 'rational' means, roughly, 'acting consistently with one's strongest normative judgement', then the connection between morality and



its demands and rational agents will be precarious at best. Morality would have a rational foundation only for those whose normative judgements (and subsequent intentions and actions) are consistent with the demands of morality. However, since this is so, the question of whether or not morality has a rational foundation *is* largely terminological. If rationality is nothing over and above consistency between one's normative judgements and one's actions, morality's supposedly inescapable and overriding normative force will not be threatened by the fact that it would be irrational for some people to act contrary to the demands of morality. If being rational is essentially being consistent, the word 'rational' is not a particularly weighty normative term.

Things look rather different if we think that the sense of 'rational' in 'a rational foundation of morality' is more like the sense given by the substantive conception of rationality. Recall that according to this conception, it is not sufficient for acting rationally that one acts consistently with one's strongest normative belief. To act rationally in the weak sense of this conception, the agent's normative beliefs must have at least some degree of warrant. On an even stronger conception of rationality the agent has to have *true* strongest normative beliefs. Someone who acts consistently, or in accordance with, his strongest normative belief and who has true beliefs can be said to be acting *fully (practically) rationally*.<sup>54</sup> An agent who satisfies these criteria can be said to be acting in full,

or complete, accordance with *the demands of practical reason*; this person is fully practically justified in doing the thing in question. To say that morality has a rational foundation, on this way of understanding things, then, is to say that it is always irrational in the strong substantive sense to fail to act in accordance with the demands of morality.

But if this is so, the interesting question becomes 'What makes a strongest normative belief true?'. This question, in turn, is equivalent to asking 'in virtue of what, if anything, is it *true* that an agent should (ought, has decisive reason to, has most reason to, etc.) do X?' If I am right about this, questions of consistency (and of rationality in the formal sense) simply aren't that important with respect to answering 'Why should I be moral?'. If we can show that it is true that we ought to act morally, then any agent who acts contrary to the demands of morality is either acting inconsistently with his strongest normative judgement (and thus irrationally on any plausible conception of rationality) or his strongest normative judgement is false. To show that morality has a rational foundation in the requisite sense, then, it will be sufficient to show that it is true that we ought to, or have most reason to, or have decisive reason to (etc.) act morally – i.e. we must show that practical reason demands that we act morally.

### 1.7 Practical reason *versus* practical reasons?

What is the connection between practical reason and practical reasons? Although the concept of a (practical) reason is important to ethics and, more broadly, to the theory of normativity, it is not clear whether it is as important as some people might think.

Although the metaphysics, semantics, and epistemology of practical reasons (and reasons more generally) are worthwhile topics to study in their own right, the central question of ethics must surely be 'What ought I do?' (a question which of course has a direct bearing on the question we are interested in). I think most people who write about ethics and who think about these issues tend to agree that this is indeed the central question in ethics. But if this is so, why is it then so fashionable to talk about reasons if the central question of ethics is concerned with how we *ought* to act? After all, the normative concept of 'a reason' is not equivalent to the normative concept 'ought' (since it could be the case that we ought to do something and have a reason for doing something else). I think the explanation is that many philosophers believe that the importance of reasons emerges once we become convinced that the concept of a reason, or its plural version, reasons, provide the fundamental building blocks of a sound theory of practical reason.

In one sense of 'practical reason' this is true enough. In so far as practical reason (or theories thereof) is concerned with what there is a reason to do, it would be very odd indeed if the concept

of a reason played little, or no role at all. But why should we think that this is what a sound theory of practical reason must consist in? We could certainly hold that answers to questions like 'How ought I live my life?' and 'What ought I do?' can only be found in practical reason, without thereby committing ourselves to thinking that these answers must be found in a correct theory of practical reasons! This would be the case only if a correct theory of practical reason simply *is* a correct theory of practical reasons. But it is far from obvious that this is the case.

To be sure, there might be a connection between the two; indeed it might even be the case that they are identical. But although it is possible that a sound theory of practical reason could simply be, or simply consist in, a sound theory of practical reasons, this must be established through argument. After all 'reason' does not mean the same as 'a reason'. Whether or not the two are necessarily connected in some way seems to me to be a substantive issue. To show that this is a substantive issue rather than a conceptual one, let me say something about the meanings of 'reason' and 'a reason' (I shall say more about these concepts in chapter four). I think it is obvious that the two notions have different meanings and that this is signalled in part by the fact that the former includes the indefinite article 'a' whereas the latter does not.

It seems clear that there is a sense of 'practical reason' in which the significance of 'a reason' (or of the plural 'reasons') is, at least on the face of it, not nearly as prominent as it would be in a theory of practical reasons. I have in mind here a sense of 'practical reason' in which 'reason' should be read with a capital 'R' so to speak. In this sense of the word, Reason is a 'faculty' (to use a rather old fashioned term) akin to, say, the faculty of sight (although there are of course obvious differences between the two). It is, as some philosophers have put it, the faculty of rational intuition.

Although it is very hard to come up with a precise definition of this sense of 'Reason', one way of characterising Reason is to view it as that which imposes constraints on our beliefs, actions, and feelings. In this respect, Reason can be compared to the law (although the law cannot meaningfully be called a faculty). Just like we are used to being told that the law demands that we do this or that, so too does Reason demand that we believe, act, or feel in such and such a way. Although there are similarities between the demands of the law and the demands of Reason, there are important differences as well; and these differences are instructive in the way of providing us with a characterisation of what Reason is. Here I shall only mention one such difference.

When we are aware of the fact that the law demands that we  $\phi$  on pain of penalty of imprisonment or monetary fines, it is still

conceptually possible that we should be justified in refusing to  $\phi$  (because of various moral and prudential considerations perhaps). However, if we are aware that Reason demands that we  $\phi$  it is not conceptually possible that we should be justified in refusing to comply with this demand. The idea is that legal demands are not necessarily justifiable in the sense that demands issued by Reason clearly are; after all, there are such things as unjust laws.

To better understand what I am talking about here, let me provide some (purported) examples of this constraint-imposing sense of 'Reason'. Perhaps the most famous is Kant's categorical imperative. Because it is so easy to overlook the constraint imposing sense of reason this has led some people to misunderstand what Kant is saying. In an unpublished paper about Kant's ethics Derek Parfit says: 'Since [the categorical imperative] is a command, it is not normative and could not be either true or false. But Kant ignores the difference between commands and normative statements, claiming that all commands are expressed with an "ought" and says that some act would be good.'<sup>55</sup> I think this comment shows that Parfit has misunderstood Kant. Although the categorical imperative is indeed an imperative, it's not just any imperative - it is demanded by Reason - or so Kant claims.

Although Parfit is correct in saying that imperatives *simpliciter* cannot be true or false (and hence not normative), Kant, to my knowledge anyway, never makes that the categorical imperative is

an imperative *simpliciter*. Compare 'The Colonel demands "stand to attention!"' and 'Reason demands "do X!"'. Both sentences are capable of being true or false. Although there is nothing normative (or at least nothing obviously normative) about the first sentence, the second sentence, if true, is as normatively loaded as you can get. A fuller version of what Kant is saying will read as follows: Reason demands that you act only on maxims which you can at the same time will to be universal law. How much more normativity can we ask for? If pure reason (whatever that turns out to be) demands that we  $\phi$ , then  $\phi$ -ing is 'as normative' as anything we could come up with.

Another example of this sense of Reason is John Broome's 'normative requirement'.<sup>56</sup> Roughly, it says that we ought to be such that we intend to do that which we believe to be a necessary means to the end we intend. Again, intending the means we believe to be necessary to our ends in this way is something Reason demands that we do; it is a constraint on our actions (or in this case, intentions) imposed by Reason.

My final example comes from John Skorupski. He says:

When I judge that  $p$ , I enter a commitment that inquirers who scrutinised the relevant evidence and argument available to them would agree that  $p$  unless I could fault their rationality or their evidence.<sup>57</sup>



What forces this commitment? Well, Reason does! It would be *unreasonable* (or, *contrary to Reason*) for an agent to judge that p while failing to fault either the rationality or the evidence of those who hold not-p.

'Practical Reason' thus construed consists of a set of demands or constraints that Reason imposes on actions. A theory of practical reason, in this sense, will correspondingly consist of some story about what those demands or constraints are, and it would be premature, *at this point*, to think that the notion of a reason will play a crucial role in this kind of theory. I think we should view Reason as being roughly that which Kant talks about in the *Groundwork* when he talks about 'Pure Reason' and 'Pure Practical Reason'. It is significant that although Kant obviously takes himself, at least in part, to be outlining a theory of practical reason, he does not, to my knowledge anyway, make any references whatsoever to the concept of a reason. Many philosophers of course believe that Kant's theory of practical reason is mistaken, but very few of these philosophers, I think, are prepared to say that the mistakes Kant made were conceptual mistakes.

Even though, according to this conception of practical reason, the notion of a reason is not obviously so important, there are other (more orthodox perhaps) conceptions of practical reason according to which the concept of a reason is central, crucial even. For



Joseph Raz 'The normativity of all that is normative consists in the way it is, or provides, or is otherwise related to reasons.'<sup>58</sup>

Similarly, Skorupski claims that

The concept of a reason seems fundamental to all thought. It is *pervasive* - actions, beliefs, and feelings (i.e. sentiments) all fall within its range; it is *primitive* - all other normative concepts are reducible to it and no other normative concept has that status; and it is *constitutive* - a being thinks if and only if it is sensitive to reasons.<sup>59</sup>

And Scanlon says:

I will take the idea of a reason as primitive. Any attempt to explain what it is to be a reason for something seems to me to lead back to the same idea: a consideration that counts in favour of it. "Counts in favor how?" one might ask. "By providing a reason for it" seems to be the only answer.<sup>60</sup>

According to this conception of practical reason (or practical normativity), since the concept of a reason is the most fundamental or even primitive concept, what we *ought* to do must be determined in some way or other by what we have most reason to do or what

we have sufficient reason to do. And what we have most reason to do, in turn, must somehow be determined by (a function of?) 'single' reasons.

## **1.8 Reasons**

In the theory of practical reason it is commonplace to distinguish between at least two kinds of reasons that are connected to (or concerned with) intentional action: explanatory reasons and normative reasons. I shall have more to say about this distinction in chapter two. As the name suggests, an explanatory reason is what we usually cite (or give) in order to explain something. Taken at face value this sort of reason is not something mysterious or problematic. The most familiar example of an explanatory reason will be that of a cause. The explanation of the fact that the grass is wet is that it rained five minutes ago; the rain caused the grass to get wet. The explanation of the fact that the Titanic sank is that it hit an iceberg; the Titanic's hitting the iceberg caused it to sink. In a similar way the explanation of the fact that a certain action took place will be provided by the cause of that action: The reason Bob went to the match is that he believed he would enjoy it. It was for this reason Bob went to the match; this is what, in the relevant sense, caused him to go.

This picture is rejected by Humeans who think that beliefs are, in themselves, never capable of motivating (and thus explaining the action of) an agent. Humeans claim that this belief must be complemented by some non-cognitive attitude (such as a desire) in order to produce a motivational state. Here the relevant desire would be the desire to do what one finds enjoyable. I will return to these Humean considerations in chapter three. Although a complete causal account of Bob's going the match will undoubtedly have to include facts about Bob's physiology and, perhaps, his surroundings, among other things, I don't think that is relevant to the present discussion. An explanatory reason, we might say, is that consideration in light of which an agent did what he did.

There are however some ambiguities we need to be aware of here. Sometimes it is tempting to ask of someone whose behaviour we find peculiar why he did what he did. If someone who is, unbeknownst to us, suffering from, say, Tourette's Syndrome suddenly starts yelling obscenities we might wonder why on earth he is doing that. What, we might ask, made him act in that way? This question is, given our ignorance of this person's affliction, a reasonable one. However, if we learn that this person is suffering from Tourette's Syndrome we should not expect to receive an explanation of his *action*. There is no action to be explained. Once we realise that the obscenities we just heard were not the product of a conscious decision to yell them out, we should conclude that in

this case the yelling of obscenities was not an action. This case is in principle no different from a person suffering from hiccups.

Although there are explanations to be given in both cases, neither of them will be explanations of *actions*. With the risk of begging serious questions in philosophy of action and mind, I think we should refer to both these cases as 'mere happenings' - not actions. On this (insufficiently argued for) account, a necessary condition of an event's being an action is that the event was consciously originated, that it was, in some sense or other, under the agent's control (whatever that turns out to mean).

I think the idea is relatively clear: hiccups, sleepwalking, yawning etc. are not actions (although, again, there is of course some explanation of why these things happen). Although there are clear examples of actions - as well as clear examples of mere happenings - it is, I believe, consistent with the position just sketched that the concept of an action is in itself vague. Consider the case of intoxication. A person who drinks one pint of lager is (usually) quite capable of performing actions. This person, when he leaves his seat at the table to order the next round at the bar, is certainly performing an action. Compare his doing so with the same person several hours later when he has consumed twelve pints. He is now so intoxicated that he literally has no control over his actions. Although his body is capable of moving (just!) he is no longer performing an action. At some point between the first and

the twelfth pint, this person stopped performing actions; at some point his bodily movements (including his slurred speech perhaps) became mere happenings. I do not intend to say anything about what a possible solution to this problem might be. Here I simply want to point out that the concept of an action (and/or agency) might be, and probably is, a vague concept: there may not be a definite moment after which a person's actions become mere happenings.

We should also take notice of another feature of explanatory reasons (or causes). The following passage is from Jonathan Dancy:

What explains why so many people buy expensive perfume at Christmas is the barrage of advertising on the television. What explains why he didn't come to the party is that he is shy. In none of these cases are we specifying considerations in light of which these things were done. But in all of them we are explaining why they were done. It seems, therefore, as if there is a wide range of things we think of as capable of giving answers to the question 'Why did he do that?' These answers range from specifying the things in the light of which the agent chose to do what he did [...] to something that is not a reason at all, really, but rather a cause....<sup>61</sup>

Unfortunately, as Dancy points out, calling those reasons that are contrasted with normative ones 'explanatory' can cause some unnecessary confusion. By contrasting normative reasons with *explanatory* reasons, the reasons thus labelled might seem to be, in some sense or other, mutually exclusive. This is not so however. First of all, there is a sense in which some, or perhaps all, normative reasons are explanatory. In this sense, when normative reasons are explanatory reasons they need not be explaining why an agent did what he did; they can be explanatory in another way. Normative reasons, can explain why, say, an act has this or that much 'going for it'.<sup>62</sup> The reason why there is something to be said for a particular act is that there is a normative reason to do it.

Sometimes (more often than not one would hope) agents act for good reason. Suppose an agent has a very strong normative reason to  $\phi$ . If the agent acts because he recognises this fact, the explanatory reason of why he did what he did will simply be his recognition of this normative reason he had.<sup>63</sup> This is not to say that normative and explanatory reasons cannot come apart. Although we can provide an explanatory reason for Bob's going to the match by citing his belief that he would enjoy it, it might be the case that Bob has no normative reason to go the match; perhaps the match (unbeknownst to Bob) has been cancelled. If the match is cancelled, then, all things being equal, there is no normative reason for Bob to go to the match. It may be tempting to think that

although Bob has a false belief about the match not being cancelled, his belief that it would be enjoyable to go to match nonetheless gives him *some* reason to, say, start walking toward the football ground. This is not so however. If we think that Bob's false belief about what he has a reason to do gives him a reason to actually do this thing, Bob's false belief would be 'bootstrapping' a reason into existence from nowhere. I shall return to, and explain the significance of, this objection later.

The normative concept of a reason (or reasons) is not specific to the practical sphere (i.e. to action). It also applies in the epistemic and emotive spheres as well. Thus we can have (good) normative reason(s) to believe that such and such is the case; we can also have (good) normative reason(s) to feel a certain way. In all three cases a certain 'reasons relation' holds between facts, agents, and actions or beliefs or feelings. Another very important feature of normative reasons is that they can be weighed against each other; and, as we shall see, this weighing dimension of reasons is sometimes said to be essential to them. In fact, some authors claim that this is in part how they should be defined.<sup>64</sup> Although an agent can have a (normative) reason to  $\phi$ , he or she could have more (normative) reason not to  $\phi$ . (Some authors make the further claim that if you have *most* reason to  $\phi$  then you *ought* to  $\phi$ . I will return to this and related issues later.)



We should also note that 'more reason' here is a bit ambiguous. There are, as I see it, two possibilities here. Suppose, all things being equal, there are two considerations (facts) such that one of them gives S a reason to  $\phi$  while the other gives S a reason not to  $\phi$ . Since we have only two reasons here, if one outweighs the other that must be because one is stronger ('weightier') than the other. On the other hand, suppose, all other things being equal, there are three considerations such that one of them gives S a reason to  $\phi$  while the other two give S reason not to  $\phi$ . Suppose further that neither of the reasons S has not to  $\phi$  is by itself capable of outweighing S's reason to  $\phi$ . If the two reasons not to  $\phi$  together are capable of outweighing the reason to  $\phi$ , then it is their *combined* weight that does the outweighing. So 'more reason' can be understood in at least two ways. In a situation where one has more reason to  $\phi$  than one has not to  $\phi$ , the weight of the reason or reasons for  $\phi$ -ing must always be greater than the weight of the reasons for doing the alternative. But this greater weight can be constituted either by one 'big' ('weighty') reason, or it can be made up of an aggregate of a number of 'smaller' (less weighty) reasons.

Of course it can also be the case that for a proposed course of action, it is not the case that there is more reason to  $\phi$  than to not- $\phi$  while at the same time it is not the case that there is more reason not to  $\phi$  than there is reason to  $\phi$ . In a case like this the reason(s) to - and not to- perform the actions are equally balanced; it is not



the case that one has more reason to perform either act. Again, this can happen in a number of ways. For example, the reason to  $\phi$  has a certain weight,  $w$  say, and the reason not to  $\phi$  also has weight  $w$ . Or, the reason to  $\phi$  has weight  $w_3$  and the two reasons not to  $\phi$  have weights  $w_2$  and  $w$ , respectively, then the combined weight of the reasons not to  $\phi$  is  $w_3$  ( $w_2 + w = w_3$ ).

I also need to say something about the relation between 'more reason' and 'most reason'. In a situation where one has only two (mutually exclusive) actions available, if one has more reason to pursue one option over the other, then one has most reason to pursue that option. Having more reason to pursue a certain option rather than another does not by itself imply that one has most reason to pursue that option. Perhaps  $S$  has most reason to pursue some third option. This could not, however, be the case in a situation where  $S$  has only two options available. In this case, having more reason to pursue one option rather than another implies that one has most reason to pursue that option.

According to some authors, whether or not an agent has more reason, as opposed to most reason, to pursue a particular course of action sometimes simply depends on how we specify the options. Suppose one specification of an agent's options is that he has three (and only three) actions ( $\phi$ ,  $\chi$ ,  $\psi$ ) available to him, and these actions are all mutually exclusive. According to these authors, another way of describing the situation would be to say that he has two options

available:  $X$ -ing and not  $X$ -ing (where  $X$ -ing just is not  $\chi$ -ing and not  $\psi$ -ing). But if we describe the situation as consisting of only two options then it does follow that if  $S$  has more reason to  $X$  than he has not to  $X$  then he has most reason to  $X$ . If, on the other hand, we describe the situation as consisting of three options, then from the fact that  $S$  has more reason to  $\phi$  than he has to  $\psi$ , we cannot infer that he has most reason to  $\phi$  - perhaps he has most reason to  $\chi$ .

If there is some controversy surrounding the question of how to weigh reasons properly (a controversy I shall return to in chapter four), the question 'when can it truly be said of an agent that he has reason to  $\phi$ ?' is even more contentious. There are, roughly, two different lines of thought on this matter: reasons are either 'internal' or 'external' (two terms to be explained presently). Since any adequate discussion about whether reasons are 'internal' or 'external' must at some stage include a discussion of Bernard Williams' famous (notorious, some would say) article 'Internal and External Reasons'<sup>65</sup>, this will be topic of the next chapter.

## 1.9 Summary and conclusion

'Why should I be moral?' is one of the oldest questions in moral philosophy, and contrary to the opinions of some philosophers, it is also a legitimate question. The kind of scepticism about morality

most commonly thought to be expressed by the familiar 'Why should I be moral?' should be seen as a scepticism about the normative status of moral considerations: are moral considerations really categorical and overriding? The task of providing an affirmative answer to these questions has traditionally been interpreted as consisting in, or involving, a demonstration that morality has a rational foundation. I have tried to show that in one sense of 'rational' this is correct. However, in the sense of 'rational' in which this is correct, the project of finding such a foundation for morality consists in nothing more(!) than a showing that moral considerations are genuinely categorical and overriding. The fact that it would, on some theories of rationality, be irrational for certain agents to act in accordance with moral demands presents no threat to the idea that moral considerations are categorical and overriding. If it can be shown that the conditional 'If  $\phi$ -ing is wrong, then there is sufficient reason to not to  $\phi$ ' really is true for all agents, then why should it matter that on some theories of rationality it would be irrational for some agents to refrain from  $\phi$ -ing? The task at hand, therefore, is to show that the conditional is indeed true. This in turn requires an investigation of when it can truly be said of an agent that he or she has a reason to, or ought to, act in a particular way. This is of course the subject matter of the theory of practical reason; it is also the subject matter of the rest of this dissertation.

---

<sup>1</sup>Kant in Guyer (ed.), p.1.

<sup>2</sup>Quoted by Geoffrey Sayre-McCord in Vallentyne (ed.), p. 105.

<sup>3</sup>Darwall (1990), p. 258.

<sup>4</sup>See Mackie, Ch. 1.

<sup>5</sup>See Williams (1985), especially Ch. 10: 'Morality, the Peculiar Institution'.

<sup>6</sup>Hume (1978), p. 283.

<sup>7</sup>See Gauthier.

<sup>8</sup>Dreier, p. 82.

<sup>9</sup>Scanlon (1996), p. 151.

<sup>10</sup>Smith, pp. 39-41.

<sup>11</sup>Miller, p. 297. For more on minimalism in general, see Wright.

<sup>12</sup>It is of course possible, though perhaps not in the end plausible, to be a cognitivist about moral discourse but not about normative discourse - and vice versa. It seems more sensible to either endorse or reject cognitivism about both domains of discourse in tandem.

<sup>13</sup>Cf. Geach, Blackburn (1986) Ch. 6, Blackburn (1988), and Hale.

<sup>14</sup>It could of course be argued that these are standards of syntax rather than standards of discipline. If this is so, then think of them simply as standards of syntax.

<sup>15</sup>See Railton, pp. 59-66. Railton also lists supervenience of the moral on the non-moral, the justifiability of moral claims, and the epistemic viability of moral claims as surface platitudes.

<sup>16</sup>Scanlon (1998), p. 63.

<sup>17</sup>I have borrowed the name 'irrealist cognitivism' from John Skorupski. For a more detailed description and defence of this position, see Skorupski (1999).

<sup>18</sup>Prichard.

<sup>19</sup>*Ibid.*, p. 1.

<sup>20</sup>*Ibid.*, p. 2.

<sup>21</sup>*Ibid.*, p. 14.

<sup>22</sup>*Ibid.*

<sup>23</sup>I shall discuss Moore's views in (slightly) more detail in chapter four.

<sup>24</sup>*Op. Cit.*, p. 8.

<sup>25</sup>Van Cleve, p. 247.

<sup>26</sup>See Goldman.

<sup>27</sup>See Nozick, pp. 172-8 and 197-227.

<sup>28</sup>*Op. Cit.*, p. 2.

<sup>29</sup>We might of course challenge this distinction. Thomas Nagel famously argued in *The Possibility of Altruism* that all reasons for action must, in the end, be viewed as agent-neutral (and thus falling into Prichard's second category). Applying Nagel's thought to the case before us, we could argue that the supposedly agent-relative reason X has for pursuing *his* happiness cannot be

distinguished from the agent-neutral reason all of us have to pursue X's happiness. In other words, we could claim that the increase in X's happiness simply *is* the good which would be 'realized either in or by the action'. (It should be pointed out that Nagel is now more pessimistic about the possibility of a reduction of agent-relative reasons into agent-neutral ones.)

<sup>30</sup>*Op. Cit.*

<sup>31</sup>Parfit (1997), p. 111.

<sup>32</sup>I shall consider the claim that all reasons for action are desire-based in the next two chapters.

<sup>33</sup>*Op. Cit.*

<sup>34</sup>*Ibid.* See p. 3.

<sup>35</sup>*Ibid.*, pp. 3-4.

<sup>36</sup>*Ibid.*, p. 4. Of course most utilitarians would object to this formulation. Although they would undoubtedly agree with Prichard that anyone's happiness is good in itself, from this it doesn't follow, on according to utilitarianism, that we ought to do whatever will promote a particular person's happiness - for instance, if the promotion of a particular person's happiness would result in a net loss of overall (or maximum) happiness. So I think we must take it that Prichard meant 'anyone' to read in plural. Furthermore, Prichard's characterisation of utilitarianism is too restrictive. Contrary to what Prichard might think, there are no obvious reasons why utilitarians (or consequentialists more generally) cannot include actions among the things that are good - there are after all such versions of utilitarianism currently on the market; see for instance Sen (2000).

<sup>37</sup>*Op. Cit.*

<sup>38</sup>Taken at face value, this claim is false. 'Oughts' apply to several normative domains, not only to the practical domain. For instance they apply also to beliefs and maybe to emotions as well, although it could of course be argued that coming to believe or coming to feel are actually *mental* actions.

<sup>39</sup>*Op. Cit.*, pp. 4-5.

<sup>40</sup>Roughly, 'satisficing' involves the idea of achieving a *satisfactory* amount of good. Thus an agent can act rightly even though he could have produced more good.

<sup>41</sup>In order to be fair to teleologists, I will substitute 'best' for 'good'.

<sup>42</sup>*Op. Cit.*, p. 4.

<sup>43</sup>*Ibid.*, p. 5.

<sup>44</sup>*Ibid.*, p. 6. The Kantian line is clear: [2.2.1] is correct, [2.2.2] is not. Kant says:

An action done from duty has its moral worth *not in the purpose* to be attained by it, but in the maxim according with which it is decided upon; it depends therefore, not on the realisation of the object of the action, but solely on the *principle of volition* in



accordance with which, irrespective of all objects of the faculty of desire, the action has been performed. (*Groundwork*, p. 399-400) Crucially, however, this does not mean that only actions that have moral worth are right.

<sup>45</sup>*Ibid.*, p. 3.

<sup>46</sup>*Ibid.*, pp. 5-6.

<sup>47</sup>*Ibid.*

<sup>48</sup>Parfit (1997).

<sup>49</sup>Rawls, pp. 48ff.

<sup>50</sup>Broome (1997), p. 139.

<sup>51</sup>Some people might not be persuaded by this example. One could for example hold that this person's 'jumping' into the canal was not really an *action* - it was merely an accident - an 'event of nature'. According to this line of thought, an *action* - as opposed to a 'mere act' (i.e. an event of nature) cannot be separated from the intention which produced it. For those who endorse this view, the proviso that the action be caused in the right way is redundant since 'the event' simply wouldn't be an action unless it was caused in the right way.

<sup>52</sup>See Broome (2001b).

<sup>53</sup>Scanlon (1998), p. 25.

<sup>54</sup>To secure that this is the case we may also have to stipulate that the fully practically rational person's true beliefs are also justified or warranted.

<sup>55</sup>Parfit (unpublished), p. 1.

<sup>56</sup>See Broome (1999).

<sup>57</sup>Skorupski (1999), p. 439.

<sup>58</sup>Raz (1999), p. 67.

<sup>59</sup>Skorupski (unpublished), p. 1.

<sup>60</sup>Scanlon (1998), p. 17.

<sup>61</sup>Dancy (2000), p. 6.

<sup>62</sup>*Ibid.*, p. 7.

<sup>63</sup>This is perhaps not as straight forward as I have here presented it. A substantial part of Dancy's latest book (2000) addresses the following worry: Suppose an agent has strong normative reasons to x. Suppose further the agent recognises this fact and acts for this reason. What is the appropriate explanatory reason here? Is it that the agent had good normative reason to x, or is it that the agent *believed* (in this case correctly) that he had good reason to x? If we choose the former explanation, it will be difficult to explain how 'normatively unsuccessful' actions come about. What if there were no good reason for the agent to x? In this is so it cannot be the case that he did x because he recognised that he had good reason to x - there was no such reason. In this case, the correct explanation seems to be that the agent merely *believed* he had good reason to x. On the other hand, if we choose the latter

---

explanation (i.e. the one proceeding via the agent's beliefs), how can an agent be said to act for good reasons if the reason for which he acted was that he *believed* he had good reason to x? This dilemma is in part forced upon us by Williams' very plausible claim that 'The difference between false and true beliefs on the agent's part cannot alter the *form* of the explanation which will be appropriate to his action.' (Williams (1981), p. 102). See Dancy (2000), especially Chs 4-8.

<sup>64</sup>See Broome (forthcoming).

<sup>65</sup>Reprinted in Williams (1981).

## CHAPTER TWO: INTERNAL REASONS

### 2.1 Introduction

This chapter is concerned with Bernard Williams' discussion and defence of what he has called 'the internal conception' of reasons for action. In his 'Internal and External Reasons'<sup>1</sup> Williams distinguishes two interpretations of the claim 'A has reason to  $\phi$ ' and argues that all true reasons statements about an agent must be given an internal interpretation. In a later article Williams says:

The central idea is that if B can say truly of A that A has reason to  $\phi$ , then...there must be a sound deliberative route to  $\phi$ -ing which starts from A's existing motivations.<sup>2</sup>

Internalism, then, is the doctrine that all reasons for action are grounded in agents' motivations. Externalism, on the other hand, denies that reasons must be connected to one's motivations in this way. Since this chapter is about internalism and not externalism (or at least not directly about externalism) I will have little to say about externalism here. For present purposes we can think of externalism as, simply, the denial of internalism.

With the risk of overlooking other, possible versions of internalism, I am simply going to assume that internalism about



reasons simply is Williams' internalism (I shall distinguish this kind of internalism from other kinds of internalisms shortly). As far as I know, it was Williams who first used the terms 'internal' and 'external' in the theory of reasons for action. Williams is still the leading internalist and anyone who is interested in defending or rejecting internalism must do so by confronting the issues first raised by Williams.

To introduce the issues we shall be considering in this chapter, let me begin by briefly addressing an issue raised by John McDowell: if all reasons for action are internal in Williams' sense, then this will bear

on a familiar problem that arises about ethical reasons in particular, in view of the evident possibility of being left cold by them. The implication of Williams' scepticism is that ethical reasons are reasons only for those for whom they are internal reasons: only for those who have motivations to which ethical considerations speak, or can be made to speak.<sup>3</sup>

Although I agree with McDowell, I will have very little to say about morality (or moral reasons) in this (and the next) chapter. Here it will suffice to acknowledge that a vindication of the apparently categorical nature of moral reasons (which, of course is our ultimate

aim) does seem to require the rejection of Williams' internalism.<sup>4</sup> On Williams' conception of reasons for action, sufficiently ruthless and sadistic amoralists would have no reason not to torture, maim, or kill others.<sup>5</sup> This is nothing new, of course. So although the potential problem highlighted by McDowell constitutes the backdrop against which the current discussion takes shape, I shall not be discussing this matter any further in this chapter. Instead of talking about morality, I want to focus on matters more internal, so to speak, to internalism.

If one wants to argue against internalism, there are, as Jonathan Dancy has recently pointed out, at least four ways of doing so: a) by counter-example; b) by showing that internalism is internally inconsistent; c) by establishing that desires are based on reasons, not reasons on desires; d) by appeal to some view in the theory of motivating reasons (triangulation).<sup>6</sup> In this (and the next) chapter I shall be employing a mixture of all four methods.

In general, the plausibility of any given theory is inversely proportionate to the number of counter-intuitive implications that theory has. If we accept Williams' account of internalism in the theory of reasons, we shall find that apart from the (perhaps?) counter-intuitive implications this theory has for morality (or moral reasons), there are several other areas in which the theory runs afoul of some of our strongly held intuitions. However, since my aim in this dissertation is to defend the idea that moral

considerations are categorical and (normatively) overriding, it would be question begging to assume that Williams account of reasons is false because moral reasons are categorical and overriding. As a result I will try to show instead that internalism is more ambiguous than people might previously have thought it to be, and that the ambiguity of internalism is not merely a superficial deficiency of the doctrine; on the contrary, the ambiguity of the various models of internalism that have so far entered the philosophical debate signals a deep problem which lies at the root of the internalist idea.

Given that internalists endorse the idea that explanatory and normative reasons are interrelated (as we shall see that Williams does), they will find it hard, if not impossible, to provide an account of reasons which on the one hand must not be grounded in counterfactual or idealised versions of what agents are or would be motivated to do, and which, on the other hand, must not merely tell us what a particular agent is currently motivated to do. In conjunction with this discussion, several other problems will also be discussed. But before we can begin to examine and assess internalism, we need to distinguish the kind of internalism this chapter is concerned with from other philosophical doctrines that are also labelled internalist.

## 2.2 Different internalisms

The following taxonomy of internalist theories in ethics is borrowed almost in its entirety from Stephen Darwall.<sup>7</sup> To begin with, we should distinguish between two kinds of internalisms: judgement internalisms and existence internalisms. The former kinds of internalisms are put forward as theories concerning the nature of our normative thought and language while the latter forms of internalisms are views about the necessary (and possibly sufficient) conditions for the existence of normative reasons. The following three internalist theses are judgement internalisms:

(1) *Morality/reasons judgement internalism*

Necessarily, if S judges (or believes, or sincerely asserts) that S is morally obliged to  $\phi$ , then S judges that S has a reason to  $\phi$ .

(2) *Moral judgement/motivation internalism*

Necessarily, if S judges (or believes, or sincerely asserts) that S is morally obliged to  $\phi$ , then S is motivated (to some degree) to  $\phi$ .

(3) *Reasons judgement/motivation internalism*

Necessarily, if S judges (or believes, or sincerely asserts) that S has a reason to  $\phi$ , then S is motivated (to some degree) to  $\phi$ .

The following theses are existence internalisms:

(4) *Morality/reasons existence internalism*

Necessarily, if S is morally obliged to  $\phi$ , then there is a reason for S to  $\phi$ .

(5) *Morality/motivation existence internalism*

Necessarily, if S is morally obliged to  $\phi$ , then S is motivated (to some degree) to  $\phi$ .

(6) *Reasons/motivation existence internalism*

Necessarily, if p is a reason for S to  $\phi$ , then S can have, and under suitable conditions would have, some motivation to  $\phi$  by virtue of a suitable awareness of p.<sup>8</sup>

I will not be discussing (1) and (2) at all, and I shall express some doubts about (3) in chapter three. (4) is of course closely related to the moral question that this dissertation seeks to address, and (5), I think, can be dismissed out of hand. It makes no sense to say

that we should be motivated by facts that we might not even be aware of. Williams' internalism should be seen a qualified version of (6), and it is this kind of internalism that this chapter will be concerned with. I take it that from (4) and (6) we can derive something like

- (7) Necessarily, if S's being morally obliged to  $\phi$  gives S a reason to  $\phi$ , then S can have, and under suitable conditions would have, some motivation to  $\phi$  by virtue of a suitable awareness of his being morally obliged to  $\phi$ .

However, on Williams' internalism, if S is aware that he is morally obliged to  $\phi$  yet fails, even under suitable conditions (as Williams sees them), to be motivated to any degree by this recognition, then S has no reason to  $\phi$ . So Williams' internalism is incompatible with (4), and this is the main reason why we need to investigate and, hopefully be able to, reject internalism.

### **2.3 Williams' Internalism**

I should begin by pointing out that Williams has returned to internalism on at least three occasions since 'Internal and External Reasons'<sup>9</sup>, and on each of those occasions Williams offers us slightly different formulations of the internalist idea. Consequently,

outlining and assessing Williams' internalism is made somewhat difficult given that there is no standard formulation of the theory's most general form. Despite differences in the formulations of internalism, one core idea remains constant:

Basically, and by definition, any model for the internal interpretation must display a relativity of the reason statement to the agent's *subjective motivational set*....<sup>10</sup>

This introduces us to a key component of internalism: the agent's subjective motivational set (S). Williams is quite happy to term all elements in an agent's S 'desires', but he is careful to remind us that 'desire' should here be interpreted in a strictly philosophical way so as to encompass such elements as 'dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may be abstractly called, embodying commitments of the agent.'<sup>11</sup> So whatever an agent has reason to do, this reason must be anchored in, derived from, or somehow related to, one or more of the elements of that agent's subjective motivational set. However, Williams says, '[w]e should not...think of S as statically given. The process of deliberation can have all sorts of effects on S, and this is a fact which a theory of internal reasons should be very happy to accommodate.'<sup>12</sup> Thus we encounter another key component of (at least some versions of)

internalism: the idea of a 'sound deliberative route'. Although this idea is somewhat indeterminate, in 'Internal and External Reasons' Williams says that sound deliberation goes beyond mere means-ends reasoning. Sound deliberation can include

...[coming] to the conclusion that one has reason to  $\phi$  because  $\phi$ -ing would be the most convenient, economical, pleasant etc. way of satisfying some element in  $S$ , and this of course is controlled by other elements in  $S$ , if not necessarily in a very clear or determinate way. But there are much wider possibilities for deliberation, such as: thinking how the satisfaction of elements in  $S$  can be combined, e.g. by time-ordering; where there is some irresolvable conflict among the elements of  $S$ , considering which one attaches most weight to (which importantly, does not imply that there is some one commodity of which they provide varying amounts); or, again, finding constitutive solutions, such as deciding what would make for an entertaining evening, granted that one wants entertainment.<sup>13</sup>

In 'Internal Reasons and the Obscurity of Blame' Williams expands the list of possible activities involved in sound deliberation:



...[deliberation] does not merely involve perceiving means to an end that has already been formulated. There are many other possibilities, such as finding a specific form for a project that has been adopted in unspecific terms. Another possibility lies in the invention of alternatives. One of the most important things deliberation does, rather than thinking of means to a fixed end, is to think of another line of conduct altogether, as when someone succeeds in breaking out of a dilemma. Yet another line of deliberative thought lies in the perception of unexpected similarities. Since there are many ways of deliberative thinking, it is not fully determinate in general, even for a given agent at a given time, what may count as 'a sound deliberative route'; and from this it follows that the question of what the agent has a reason to do is itself not fully determinate.<sup>14</sup>

Williams also builds into his conception of deliberation that every agent has a 'general interest in being factually and rationally correctly informed.'<sup>15</sup> Thus Williams' internalist can (rightly) deny that desires based on false beliefs give rise to reasons. Consider the case he presents us with:

The agent believes that this stuff is gin, when it is in fact petrol. He wants a gin and tonic. Has he reason, or a reason, to mix this stuff with tonic and drink it? ...it is just very odd to say that he has a reason to drink this stuff, and natural to say that he has no reason to drink it, although he thinks he has.<sup>16</sup>

For Williams, then, it is not only odd to say that this person has a reason to drink the stuff before him, it is false. And the reason why it is natural (and correct) to say that the agent has no reason to drink the stuff before him is that this desire is prompted by a false belief: the belief that the stuff before him is gin. However, and this will turn out be very important, if the agent does drink the stuff we could then correctly cite the reason why he did it. This brings out the *explanatory* dimension of reasons: the reason he drank the stuff was that he believed it to be gin and he wanted a gin and tonic. However, the sense in which it is natural to say that he has no reason to drink the stuff is the *normative* sense. And it is this sense of 'reason' Williams' paper is mainly about. Crucially, however, Williams claims that there is an important relation between these two dimensions:

It must be a mistake simply to separate explanatory and normative reasons. If it is true that A has a reason to  $\phi$ ,

then it must be possible that he should  $\phi$  for that reason;  
 and if he does act for that reason, then that reason will be  
 the explanation of his acting.<sup>17</sup>

Let us call this the 'interrelation thesis' or (IT) for short. In  
 'Internalism and the Obscurity of Blame', Williams claims that the  
 truth of (IT) provides one of the fundamental motivations behind  
 the internalist account.<sup>18</sup> This is a very important principle and we  
 shall have good cause to return to it later. Let me end this  
 summary of Williams' internalism by listing four statements that  
 are, according to Williams, true of internalism:

- (i) An internal reason statement is falsified by the absence of some appropriate element from S.
- (ii) A member of S, D, will not give A a reason to  $\phi$  if either the existence of D is dependent on false belief, or A's belief in the relevance of  $\phi$ -ing to the satisfaction of D is false.
- (iii) (a) A may falsely believe an internal reason statement about himself, and (we can add)  
 (b) A may not know some true internal reason statement about himself.
- (iv) Internal reason statements can be discovered in deliberative reasoning.<sup>19</sup>

This brings my summary of Williams' internalism to an end.  
 All that remains for me to do is to state, in its most general form,

the theory itself. As I said earlier, since Williams has produced various versions of this general form over the years, one should perhaps be careful in stating *the* general form of internalism. However, given the information we have assembled throughout this summary, I believe the version of internalism that best captures these elements is the one Williams puts forward in 'Internalism and the Obscurity of Blame'. Internalism about (practical) reasons for action says:

(IR)        A has reason to  $\phi$  if and only if A could reach the conclusion to  $\phi$  by a sound deliberative route from the motivations A already has.<sup>20</sup>

However, before we can proceed to critically investigate whether this account of reasons is plausible, there is an important ambiguity in (IR) that we need to attend to first.

## 2.4 Disambiguating Internalism

We should begin by noting that although (IR) provides us with what Williams takes to be the necessary and sufficient conditions for its being true that someone has a reason to do something, Williams sometimes writes as if thinks that (IR) reveals more than this. At one point he asks 'What are we *saying* when we say someone has a

reason to do something?'<sup>21</sup> Similarly, he says "A has reason to  $\phi$ ' *means* more than 'A is presently disposed to  $\phi$ '<sup>22</sup>; and 'I think the *sense* of a statement of the form 'A has reason to  $\phi$ ' is given by the internalist model.'<sup>23</sup> These quotations strongly suggest that Williams is, perhaps implicitly, trying to explain the *meaning* of reasons statements.<sup>24</sup>

If we take (IR) to be (at least a partial) analysis of the meaning of 'A has reason to  $\phi$ ', the analysis is hampered by the fact that the analysans is grammatically flawed. Since ' $\phi$ ' stands for some verb of action, the wording of the analysans does not sit well with our ordinary use of English. In English the verb 'conclude' always takes on a noun clause and not an infinitive clause as in Williams' sentence. Thus we say 'he could conclude *that ...*' So why is this important? Well, since English grammar does not permit this kind of formulation, it is hard to understand exactly what Williams' formula is supposed to mean. And this is of course an especially unwelcome result if (IR) is (in part) suppose to reveal the meaning of 'A has reason to  $\phi$ '.<sup>25</sup> No problem, we might think, let's just substitute an unproblematic phrase for the problematic 'reach the conclusion to'. Here, however, problems arise. What do we put in its place? Here are a few suggestions:

- (i) 'conclude that A has reason to'

If this is what the phrase 'reach the conclusion to' meant, (IR), although not perhaps implausible with regards to the necessary and

sufficient conditions for someone's having a reason, would be useless from the point of meaning analysis. The analysans cannot make use of a term appearing in the analysandum. This should be fairly obvious; we cannot use the idea of a reason to explicate the idea of a reason. Nor can we replace the phrase with

(ii) 'conclude that A ought to (should)'

Williams actually suggests this reading at one point.<sup>26</sup> If this is what (IR) means, (IR) is almost certainly false. It would be too restrictive. A could very well have a reason to do something without it being the case that A could conclude that he ought to do that thing. In one sense of 'could' it is arguably true that for any reason A has, that reason could have been his strongest reason, and that, consequently, A could conclude that he ought to act on that reason. But this would be an odd condition to impose on an agent's having a reason. Why should it be a necessary condition for the obtaining of a reason that that reason could be stronger than it is? Here is yet another suggestion:

(iii) 'decide to'.

Although this suggestion might initially seem plausible, it should be rejected on the same grounds as (ii). A cannot (rationally) *decide* to go for a walk *and* stay at home at the same time. The problem with (ii) and (iii) is that they cannot adequately account for the existence of reasons that pull in different directions so to speak (except of course in that limited sense of 'could' I mentioned

earlier). But this, of course, is something an adequate theory of reasons must be able to do. Furthermore, anyone who accepts (ii) or (iii) will be unable to account for, or make sense, of 'outweighed' or 'defeated' reasons. I will have more to say about these kinds of reasons and internalism's treatment of them in subsequent sections. The final, and, I believe, most plausible suggestion is

(iv) 'be motivated to'

I suspect that this is essentially what Williams takes IR to mean; although, as we shall see, IR can be interpreted in a way that does not make any reference to what the agent would *actually* be motivated to do. The current proposal, though, runs as follows:

A has reason to  $\phi$  if and only if A could be motivated to  $\phi$  by a sound deliberative route from the motivations A already has.

Unfortunately, this formulation is not entirely satisfactory either. Sound deliberative routes do not, in and of themselves, motivate; nor do they lead agents to become motivated to, say,  $\phi$ . However, sound *deliberation* might. Obviously, what the internalist needs to capture here is the idea that motivation can result from *engaging* in sound deliberative reasoning. However, it is not clear whether A, himself, has to engage, or be able to engage in this type of reasoning in order for him to have the reason(s) in question. It might be enough for it to be true of A that he has reason to  $\phi$  that a



sound deliberator (other than himself) who has full knowledge of A's S informs him that 'there is something to be said for  $\phi$ -ing' given his S. Suppose for instance that A has a motive that would be served by his giving the correct answer to a multiple choice question about a difficult mathematical problem. Suppose the right answer is 'c' but A is not able to figure this out for himself - perhaps because he is very bad at maths. His friend, who is a good mathematician (and who recognises that 'c' is the correct answer) and who has knowledge of (at least some of) the elements of A's S, could nonetheless correctly say of A that he has a reason to pick 'c'. As Williams puts it, 'What we can correctly ascribe to him in a third-personal internal reason statement is also what he can ascribe to himself as a result of deliberation...'.<sup>27</sup> Keeping this 'third-person proviso' in mind, my final suggestion for the best formulation of internalism (at this stage) is:

- (IR\*)      A has reason to  $\phi$  if and only if A could become motivated (to some degree) to  $\phi$  as the result of having engaged in sound deliberation on, and from, the motivations he already has.

## 2.5 Two Further Distinctions

The first thing we should note about (IR\*) is that it too is ambiguous. What does it mean to say of someone that he or she could be motivated to  $\phi$ ? As far as I can see, there are two ways of interpreting this claim. On the one hand it can be read as a psychological claim about what A *would*, as a matter of fact, be motivated to do if he deliberated correctly. I think this interpretation is the most natural one. According to this interpretation, there are two individually necessary and jointly sufficient conditions for its being true that A has a reason to  $\phi$ , namely:

- 1) A has some motive in his S which would served by his  $\phi$ -ing.
- and
- 2) If A were to become aware of (1) he *would* be motivated to  $\phi$ .

It is important to keep in mind that A could well be motivated to  $\phi$  without having acquired that motivation *through* deliberation. A's motivation to  $\phi$  could simply be a *sui generis* (so to speak) element in his S - a 'basic' desire. I shall call this view *actualism*.

- (A)            A has reason to  $\phi$  if and only if A *would* be motivated to  $\phi$  after having deliberated soundly on, and from, the motivations A already has.

According to the other interpretation, what we need to determine is whether A *could* be motivated to  $\phi$  after having deliberated soundly from his S; whether or not A will *in fact* be motivated to  $\phi$  is irrelevant to its being true that A has a reason to  $\phi$ . According to this interpretation, the possibility of A's being motivated to  $\phi$  is, irrespectively of what he would actually be motivated to do, guaranteed by the existence of some motive in his S which is such that it would be satisfied or promoted by his  $\phi$ -ing. On this interpretation, it will be necessary and sufficient for A's having a reason to  $\phi$  that he has some motive that would be satisfied by his  $\phi$ -ing. Let us call this view *possibilism*.

- (P)            A has reason to  $\phi$  if and only if A has some motive which will be served by his  $\phi$ -ing.

Both actualism and possibilism thus hold that (1) is a necessary condition for A's having a reason to  $\phi$ . Possibilism makes the further claim that (1) is sufficient as well. Actualism denies this.

So which approach does Williams favour? Is he an actualist or a possibilist? Answering this question would require a level of exegetical exposition which would take us too far afield for present purposes. Since it is not crucial for the rest of my argument, I shall not try to establish whether Williams is an actualist or a possibilist -

although, I suspect that he would be more inclined to endorse actualism. Let us look at possibilism first.

## 2.6 Possibilism and motives

The possibilist wants to say that although (IR\*) is true, it is potentially misleading. What the modal locution 'could be motivated to' in the right-hand side of the biconditional *really* means is simply '*there is* some motive in A's S that would be served by his  $\phi$ -ing - full stop!'<sup>28</sup> However, an agent can have a motive that either consists in, or which would be served by, his  $\phi$ -ing without being motivated to  $\phi$ . (And, we might add, an agent can be motivated to  $\phi$  without its being the case that  $\phi$ -ing serves some motive of his. Just like in Williams' example of the would-be petrol drinker, an agent can be motivated to do something because he falsely believes that doing this thing will serve a motive of his.) So the possibilist will stress the importance of separating motives from motivations. So what is a motive? The most natural suggestion is that motives simply are desires. When we say that A has a motive which would be served by his  $\phi$ -ing, we normally mean that A has a desire which would be satisfied by his so doing.

However, although desires are obviously members of an agent's motivational set (and are thus motives), this does not imply that all elements of an agent's motivational set (i.e. his motives)

either are, or can be reduced to, desires. As we shall see in the next chapter, Williams sometimes writes as if he thinks that beliefs can be elements of an agent's motivational set as well.<sup>29</sup> However, for convenience I shall simply assume that all elements in an agent's *S* can be reduced to a desire or a desire-like state. I think it is fairly safe to call the elements of an agent's *S* 'desires' as long as we keep in mind

that *S* can contain such things as dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may be abstractly called, embodying commitments of the agent.<sup>30</sup>

So how should we understand these elements? I think we can begin by making a distinction between two broad kinds of desires: occurrent and non-occurrent desires. To have an occurrent desire to  $\phi$  (i.e. to be aware of the fact that one desires to  $\phi$ ) involves, at least in part, experiencing certain psychological feelings, analogous to bodily sensations, towards  $\phi$ -ing.<sup>31</sup>

The connection between having an occurrent desire and experiencing certain psychological feelings is not, I think, merely contingent in the sense that we could have the former while lacking the latter. In other words, the having of an occurrent desire is necessarily accompanied by a certain phenomenology - to have an

occurrent desire is, in part, to *feel* a propensity or aversion towards that which one desires to do or to avoid. But this, of course, is nothing other than being motivated to do or avoid the thing in question.

This should be contrasted with the having of a non-occurrent desire which necessarily lacks these phenomenological properties (because if it did not, then it would simply be an occurrent desire). But if non-occurrent desires necessarily lack these phenomenological properties, what licenses us to call these things desires in the first place? It might be objected that non-occurrent desires simply aren't desires at all. According to what Michael Smith has called 'the strong phenomenological conception' of desires, 'desires are, like sensations, simply and essentially states that have a certain phenomenological content.'<sup>32</sup> I agree with Smith that this conception of desires is mistaken. There is a clear sense in which A can have a desire to  $\phi$  without feeling a propensity toward  $\phi$ -ing in exactly the same way someone can have a motive that would be served by  $\phi$ -ing without being motivated to  $\phi$ . If we reflect on the matter, I think we should all agree that we retain our desires even when we are not experiencing them. It certainly seems correct to say that we retain our desires even when we are asleep, absent-minded, or hard at work on some project that requires our undivided attention. Even if we don't experience certain desires under these circumstances, this does not mean that

the desires aren't there; they are, but they're dormant, we might say. But being dormant just is being non-occurrent.

On the current (non-phenomenological) proposal, then, what does it mean to say of A that he has a desire to  $\phi$ ? Well, since we can have a desire without experiencing certain psychological feelings, the experiencing of these psychological feeling cannot be the distinguishing feature of desires as such. I think we should understand the claim that A has a desire to  $\phi$  as the claim that A is disposed to act in a certain way under certain conditions. Thus A's desire to eat an ice cream consists in its being the case that A would eat an ice cream in circumstance C. The specification of C will quite often have to include counterfactuals about the absence of stronger desires and, in this case, the availability of ice cream among other things. Thus to say of A that he desires to eat an ice cream is to say that were it the case that A had no stronger desires to the contrary and were it the case that there were some ice cream about, then A would eat it. We can represent this thesis as

- (D)        A's desire to  $\phi$  consists in A's being disposed to  $\phi$  in circumstance C (where the specification of C contains the relevant counterfactual conditions).



A dispositional account of desires also allows us to make sense of the idea that some desires are accompanied by a feeling of propensity toward the object of the desire. As Smith puts it:

...according to this conception, desires have phenomenological content just to the extent that the having of certain feelings is one of the things that they are dispositions to produce under certain conditions. Some desires may be dispositions to have certain feelings under all conditions: these have phenomenological content essentially. Other desires, though they are dispositions to behave in certain ways, may not be dispositions to have certain feelings at all: these lack phenomenological content altogether.<sup>33</sup>

Although this is a fairly rough and ready account of what it means for someone to have a desire, I think it is adequate for my purposes here.

I should point out that by accepting this theory of desires, I am not thereby accepting Humeanism about motivation. I have said that desires are dispositions, but this does not mean that only desires dispose us to act. Beliefs may be capable of doing so as well. But since this is so, the counterfactual conditions mentioned in (D) may have to include such conditions as 'A does not believe

that he ought not  $\phi'$  etc. Given this analysis of desires (which I think we should accept), it turns out that (P) is not easily reconciled with certain other internalist theses.

## 2.7 Counterexamples to possibilism

Recall that Williams says that a necessary condition for the obtaining of a reason for action for A is that this reason is potentially explanatory of A's action:

(IT):        If A has a reason to  $\phi$ , then it must be possible that A should  $\phi$  for that reason.<sup>34</sup>

Williams says that (IT) expresses an important truth about normative reasons, and that the truth of (IT) presents a serious challenge to the external reasons theorist. As I said earlier, Williams actually goes so far as to say that (IT) is one of the 'fundamental motivations of the internalist account.'<sup>35</sup> The idea, spelled out in more detail, is this:

It must be a mistake simply to separate explanatory and normative reasons. If it is true that A has a reason to  $\phi$ , then it must be possible that he should  $\phi$  for that reason; and if he does act for that reason, then that reason will be

the explanation of his acting. So the claim that he has a reason to  $\phi$  - that is, the normative statement 'He has a reason to  $\phi$ ' - introduces the possibility of that reason being an explanation; namely, if the agent accepts that claim (more precisely, if he accepts that he has more reason to  $\phi$  than to do anything else). This is a basic connection. When the reason is an explanation of his action, then of course it will be, in some form, in his S, because certainly - and nobody denies this - what he actually does has to be explained by his S.

So why does this push us in the direction of internalism? Derek Parfit has suggested that Williams may have the following argument in mind:

- (1) Normative reasons must be able to be motivating reasons. It must be possible that we should act *for these reasons*.
- (2) Motivating reasons must be internal, since our acts must be in part explained by our desires, or other motivating states.

Therefore,

- (3) Normative reasons must be internal.

Although Parfit acknowledges that Williams may not have intended this argument, he says, correctly in my mind, that this argument is suggested by Williams' remarks in Williams (1981), pp. 102 and 106-7, and in Williams (1995), p. 39 (see Parfit, p. 112).

Regardless of whether Williams intended this argument or not, it should be rejected because it is invalid. What follows from (1) and (2) is not (3), but

(3\*) Normative reasons must *be able to be* internal.

I can see no obvious reason why externalists could not accept (3\*). The distinctive externalist claim, as I understand it, is not that true reasons statements are in some way or other 'potentially evidence transcendent' (which may be the central claim of certain kinds of *realism*); rather, externalists reject the claim that true reasons statements are *made true* by being related to various features of our motivational sets.

Now A's having a reason to  $\phi$  can never by itself be sufficient to explain his  $\phi$ -ing if he  $\phi$ 's. As part of his argument against the external reasons theorist, Williams says:

...nothing can explain an agent's (intentional) actions except something that motivates him so to act. So something else is needed besides the truth of the external

reason statement to explain action, some psychological link; and that psychological link would seem to be belief.<sup>36</sup>

But this point is equally applicable to internal reasons statements. Since, as Williams puts it, 'A may not know some true internal reason statement about himself'<sup>37</sup> (and, we may suppose, this lack of knowledge is attributable to the fact that A has no beliefs about the reason statement in question - as opposed to believing it but not being justified in believing it), it would be incredible if internal reason statements could, by themselves, do all the explaining. So the internalist must also accept that a 'psychological link' between the true reason statement and the agent must figure in a correct explanation of intentional action.<sup>38</sup> There is also another problem with the idea that 'A has reason to  $\phi$ ' introduces the possibility of A's acting for that reason. In non-philosophical, everyday explanations of actions, it is perfectly fine to say that the reason why A did what he did was that he believed that he had a (good) reason to do it, but from a philosophical point of view this kind of explanation is at best incomplete. A more philosophically robust explanation must also take into account that when A  $\phi$ 's intentionally he does so, as Williams puts it,

not because he believes only that there is some reason or other for him to  $\phi$ , but because he believes of some

determinate consideration that it constitutes a reason for him to  $\phi$ .<sup>39</sup>

I think Williams is right here and it seems that this point can be generalised to cover not only action explanations but reasons statements as well. Thus if it is true that A has a reason to  $\phi$ , then there must be some consideration,  $p$ , in virtue of which he has a reason to  $\phi$ . This much should be admitted by everyone, externalists included. If we apply this consideration to the interrelatedness of explanatory and normative reasons, we get the following picture. If the fact that the tickets are selling out fast is a reason for A to get on the phone now, then it must be possible that A should get on the phone now because the fact that the tickets are selling out fast gives him a reason to do so. In other words, it must be possible that A's getting on the phone now could be explained by the fact that the tickets are selling out fast and that this gives him a reason to do so. Thus we get

(IT\*):        If the fact that  $p$  gives A a reason to  $\phi$ , then it must be possible that A should  $\phi$  because the fact that  $p$  gives A a reason to  $\phi$ .

But, as we have seen, A can  $\phi$  because the fact that  $p$  gives him a reason to do so only if he believes that, or is aware of, the fact that

p gives him a reason to  $\phi$ . Again, there is no reason for the externalist to disagree with any of this. This will turn out to be of some significance later.

Consider next the following case: A has with great enthusiasm and excitement witnessed the pomp and circumstance of various military parades on television. He too wants to parade around with medals, in uniform, in front of the Queen. Call the content of this desire ' $\phi$ '. Now, suppose that in order for A to parade in front of the Queen he must complete a certain amount of military training, and that a necessary component of his military training consists in his going through boot camp (call this ' $\psi$ '). Thus  $\psi$ -ing is a necessary means to satisfy the desire to  $\phi$ . According to (P), A has reason to  $\psi$  (i.e. A has a motive which would be promoted by his  $\psi$ -ing). Boot camp, however, is not an enjoyable experience. Suppose that when A finds out that he has a reason to  $\psi$ , this causes A to lose his desire to  $\phi$ . Assuming that A has a reason to  $\psi$  only because of the relation between  $\psi$ -ing and his desire to  $\phi$ , then since A no longer desires to  $\phi$  he has no reason to  $\psi$ . In other words, A's recognition that he had a reason to  $\psi$  contributed to making it the case that he has no reason to  $\psi$ .

This little story, I think, shows that possibilism is at least *prima facie* incompatible with (IT). According to (IT), if A has a reason to  $\phi$ , then it must be possible that he should  $\phi$  for that

reason. But in the example I have just described this does not seem to be possible. Since it was A's recognition of the fact that [his going through boot camp is a necessary means for his parading in front of the Queen] that caused him to lose the desire to parade in front of the Queen, the (supposed) reason A had to go to boot camp could never have been the explanation of why A went to boot camp, and hence it could not have been a reason for A to go through boot camp.

Of course the possibilist could say that certain reasons (like certain dispositions perhaps) are 'finkish' - i.e. that it was true of A that he *did* have reason to go through boot camp up until the point just before he recognised that he had a reason to do so. But his recognition of this fact makes it the case that he (now) has no reason to go through boot camp.<sup>40</sup> But why would this help the possibilist? Finkish reasons, if there are any, would have to be defined, at least in part, *by their failure to meet the (IT) criterion*: finkish reasons are such that they could never figure in a correct explanation of A's acting *for that reason*. And if that (supposed) reason could never have explained A's action, it follows according to (IT) that it was never a reason in the first place. The existence of finkish reasons is incompatible with (IT).

The possibilist might dig in his heels and respond by saying that A really does have a finkish reason to  $\phi$  and that finkish reasons are not a threat to (IT). He might argue that what is



important here is not whether A *would* act for that reason, rather, what is important is whether it is possible that he *could* act for that reason. In other words, the possibilist could maintain that there are finkish reasons, and, more importantly, that finkish reasons do not provide a counterexample to (IT) - *it is possible* that A should go through boot camp because he recognises that he has reason to do so; the counterfactual isn't impossible, it just wouldn't happen because of the contingent make up of A's motivational set. The idea is this. Had A retained his desire to parade in front of the queen after he found out what this would require of him, the reason to go through boot camp would have been sustained after deliberation. And surely, the thought goes, this *is* possible. And if this is possible it is also possible that A should have gone through boot camp because he recognised that he had reason to do so. Thus he *could* have acted for that reason; and therefore the existence of this reason does not violate (IT).

I don't think this is right. In one sense *it is possible* that A should go through boot camp - there is a possible world in which A would do so. However, I don't think this move is open to the possibilist, or indeed to any internalist. In his reply to John McDowell's suggestion<sup>41</sup> that what an agent has reason to do is whatever an Aristotelian *phronimos* would do, Williams says:

...in considering what A has reason to do, one thing that A should take into account, if he is grown up and has some sense, are the ways in which he relevantly fails to be a *phronimos*. Aristotle's *phronimos*...was, for instance, supposed to display temperance, a moderate equilibrium of the passions which did not even require the emergency semi-virtue of self-control. But if I know that I fall short of temperance and am unreliable with respect even to some kinds of self-control, I shall have good reason not to do some things that a temperate person could properly and safely do.<sup>42</sup>

This passage suggests that internalists (of any persuasion) should be very reluctant to ground an agent's reasons in counterfactually idealised version of that agent or his S.

I think internalists are committed to holding that what is possible and impossible in the present context must be determined or controlled by A's *actual* S. The elements in A's actual S (including his dispositions) limit the range of possible action explanations, and this in turn limits the range of reasons that apply *to him*. Of course, it is logically possible that A's S could have been different (i.e. it could have been such that A retained his desire to parade in front of the queen), but since there are no restrictions on what sort of S A could have had (except certain logical restrictions),

there are no restrictions on which action explanations are possible, and, consequently, no restrictions on what could be a reason for A. But if there are no restrictions on what could count as a reason for A, (IT) becomes more or less vacuous - nothing, except certain logically impossible actions, would be ruled out by it. This is of course an unwelcome conclusion for the possibilist as he wants to say that (IT) is not vacuous - that it is an important principle with real cutting-power which provides a 'fundamental motivation behind the internalist account of reasons'. To avoid this conclusion, then, possibilists (and internalists more generally) must stress the necessity of grounding the possible reasons an agent could have in that agent's actual S. I am going to call this the 'actualist constraint':

- (AC)        What is possible and impossible *with respect to explanations* of A's actions is determined by A's *actual* S.

Let us be clear about a few things here. I said earlier that an agent's having a desire must be analysed in terms of certain counterfactuals, including counterfactuals about the agent's S itself. On this analysis, A's desire to eat an ice cream consists in his disposition to eat ice cream in circumstance C, where C is specified in terms of its being the case that were A to have no stronger

contrary desires and were there some ice cream around, A would eat it. But this analysis seems to be incompatible with (AC). How can we retain (D) after we have introduced (AC)? Suppose A desires to eat ice cream but that he also has a stronger desire to do something other than eating ice cream, would the analysis of A's desire to eat an ice cream then not involve the sort of counterfactual analysis of A's S that (AC) says is illegitimate? The answer is no. The counterfactual analysis we looked at earlier, (D), is an analysis of an agent's having *a desire*. (AC), on the other hand, is a constraint on what could be an explanation of A's action. It should be clear that we need two different analyses here; one for desires and one for explanations of actions. Unlike reasons or action explanations (given (IT) and (AC)), it seems that desires can be finkish. That this is a genuine possibility should come as no surprise since the phenomenon of finkishness is usually associated with dispositions (whether it be dispositions of inanimate objects or agents) and we have provided a dispositional account of desires. Suppose A's desire to  $\phi$  is such that whenever he becomes aware of the fact that he has that desire that very awareness produces in him an even stronger desire to avoid  $\phi$ -ing. This would be an example (I think) of a genuinely finkish desire given our analysis of desire. However, even in the light of the possibility of there being such desires, we can still retain our original analysis of desires. A's desire to  $\phi$  consists in his disposition to  $\phi$  in circumstance C (where

A has no stronger desires to the contrary). Since it is possible that circumstance C should obtain (i.e. there is a possible world in which circumstance C obtains), it is also possible that A should  $\phi$ . As before, the counterfactual is not impossible, it's just that it would never be realised in the actual world (because in the actual world, A is such that whenever he becomes aware of the fact that he desires to  $\phi$  he desires even more strongly not to  $\phi$ ). So although it is possible that A should  $\phi$ , this sense of 'possible' is of little or no use to Williams and his fellow internalists. Williams needs a sense of 'possible' that is stronger than the one we have been employing so far.

(AC) says that the modal term 'possible' must be restricted in such a way that it does not apply to the agent's S itself. The agent's actual S must be fixed. The internalist must insist that we think of the relevant possibilities in terms of A's actual S, not in terms of an S that A could have had. So if (IT) is constrained by (AC) we can reformulate (IT) (and thus (IT\*)) by substituting the word 'feasible' for 'possible':

(IT\*f): If the fact that p gives A a reason to  $\phi$ , then it must be *feasible* that A should  $\phi$  for the reason that the fact that p gives him reason to  $\phi$ .<sup>43</sup>

It simply cannot be the case that a refutation of (IT\*) requires a showing that

The fact that  $p$  gives  $A$  a reason to  $\phi$  but it is not possible that  $A$  should  $\phi$  because the fact that  $p$  gives  $A$  a reason to  $\phi$ .

Nobody could meet this requirement since this would require us to show that

It is not possible that  $A$  should  $\phi$  because the fact that  $p$  gives  $A$  a reason to  $\phi$ .

But how could we show this? It simply cannot be a necessary truth that  $A$  should fail to  $\phi$  because the fact that  $p$  gives  $A$  a reason to  $\phi$ . But if it is impossible to show that the consequent of (IT\*) is false, then (IT\*) is irrefutable. However, consider the contraposed version of (IT\*):

(IT\*<sub>C</sub>): If it is not possible that  $A$  should  $\phi$  because the fact that  $p$  gives  $A$  a reason to  $\phi$ , then it is not the case that the fact that  $p$  gives  $A$  a reason to  $\phi$ .

I think it is fairly clear that Williams wants to be able to derive the conclusion

It is not the case that the fact that  $p$  gives  $A$  a reason to  $\phi$

from  $(IT^*_C)$  and the proposition

It is not possible that  $A$  should  $\phi$  because the fact that  $p$  gives  $A$  a reason to  $\phi$ .

But this derivation would also require a showing that it is impossible that  $A$  should  $\phi$  because the fact that  $p$  gives  $A$  a reason to  $\phi$ ; which, again, is something that cannot be shown. For this reason, then, I think Williams intends the word 'possible' not to be read as 'metaphysically possible'. I have suggested that we introduce 'feasible' to denote this weakened sense of 'possible'. How we should interpret 'feasible' I shall leave unexamined.

Suppose next that  $A$  has, in his actual  $S$ , a disposition to refrain from doing that which he believes involves getting sweaty and dirty, and that he believes (plausibly enough) that going through boot camp entails that he will get sweaty and dirty. In this case, the combination of  $A$ 's disposition and his belief will prevent him from going through boot camp. So now we have a disposition in  $A$ 's actual  $S$  which would (along with the relevant beliefs) prevent

him from acting on the supposed reason he has to go through boot camp. Importantly, however, this does not mean that A could not retain his desire to parade in front of the queen after deliberating about the necessary means to satisfy that desire. Whether he does or does not retain his desire to parade in front of the queen is a contingent psychological matter which is presumably going to be determined by other features in his S and (perhaps) various other features that make up his over all psychology.

The problem for the possibilist, then, is this. If A has a disposition of such a kind that it prevents him from acting in a way that he believes entails him getting dirty, then although A might have a motive which would be served by his performing an action that will result in his getting dirty, it is not feasible (i.e. possible - in the relevant sense) that A should go through boot camp, and therefore it cannot be the case that A has a reason to go through boot camp. But this is incompatible with (P). According to (P), A *does* have reason to go through boot camp. How can the possibilist respond to this difficulty? Should he perhaps suggest that A's desire to parade in front of the queen gives A a reason to try to get rid of his disposition? or that A's desire gives him reason to acquire certain false beliefs about what going through boot camp involves perhaps? Questions multiply....

The problem with possibilism is the right-to-left implication in (P). Given (IT) and (AC) it is simply false that A's having a motive



which will be served by his  $\phi$ -ing is sufficient for A to have a reason to  $\phi$ . The opposite implication, from left-to-right, does however seem to capture an important element of the internalist doctrine. But what is the argument for this implication? The only one I know of is the one suggested by Parfit on Williams' behalf. As I have said, we need not accept this argument.

To get round this problem, perhaps the possibilist should instead focus on the fact that although A's desire to parade in front of the Queen was not formed on the basis of a false belief, his desire to parade in front of the Queen was formed on the basis of incomplete information. The possibilist needs to take a more holistic approach to specifying what reasons statements are true given (AC) and (IT). As Williams suggests: had A been better informed, 'he may [have] come to have some more concrete sense of what would be involved, and [as a result] lose his desire for it...'. In other words, had A known the relevant facts (such as  $(\phi \rightarrow \psi)$ ) he might not have desired to  $\phi$  in the first place.

The obvious question for the possibilist is 'how do we specify the 'relevant facts'?' Clearly he must be able to rule out facts that are irrelevant to a particular deliberative situation. If we are investigating, say, whether A has a reason to go to the cinema tonight, it would be very odd indeed if the fact that [Ulan Bator is the capital of Mongolia] should play a role in determining whether A has this reason or not. So by what criteria could the possibilist sort

the irrelevant facts from the relevant ones? It seems to me that the only way to do this would be to say that the relevant facts are those facts, the awareness of which, would affect, or make a difference to what A would (actually) be motivated to do. But this, of course, is nothing short of actualism. In the end, then, I think that in light of these considerations the internalist will do better (for the time being anyway) by rejecting possibilism in favour of actualism.

## 2.8 Actualism and accidie

According to actualism, then, the two (individually necessary) conditions

- (1) A has some motive which will be served by his  $\phi$ -ing; and
- (2) If A were aware of (1) he *would* be motivated to  $\phi$ .

together constitute a sufficient condition for its being the case that A has a reason to  $\phi$ . So, taking this approach we must ask 'would A, given his current psychological and motivational make up, after sound deliberation, actually be motivated to  $\phi$ ?' If the answer is no, then, even though sound deliberation could reveal that some motive of his would be served by his  $\phi$ -ing, it is not true of A that he has a reason to  $\phi$ .

I think there are several problems with this view, the most serious of which is presented by the phenomenon of accidie. Before I discuss accidie, I want to briefly mention two other cases in which we clearly want to say that an agent has a reason to do something although he is, in the relevant sense, incapable of doing it.

(i) Buridan's ass cases

Suppose A has deliberated soundly on all relevant facts concerning the actions currently available to him and as a result he is equally motivated to  $\phi$  and to  $\psi$  (where it is not possible to do both) and there is nothing else he is more motivated to do. In other words, A is equally most motivated to  $\phi$  and to  $\psi$ . Being a staunch internalist, this leads A to believe that he has most reason to  $\phi$  and he has most reason to  $\psi$ . Finally, suppose also that A has a disposition, (R), such that he never acts contrary to what he believes himself to have most reason to do (i.e. he is rational in at least the formal sense discussed in section 1.6). A thus finds himself in the company of Buridan's ass, being equally most motivated to perform two incompatible actions. But if he should find himself in this unfortunate situation where his motivations are exactly balanced and if his actual S contains this particular disposition, then he will be incapable of performing either action. But if this is so, then neither reason is potentially (in the relevant

sense) explanatory, and hence he has no reason to perform either action.

In the literature, the case of Buridan's ass is usually presented as a case in which the unfortunate animal's reasons are exactly balanced and it starves to death as a result of being rational. Most commentators suggest that the moral of that particular story is that we should not accept that rationality consists simply in being disposed to act in accordance with one's (first-order) normative beliefs. Whether or not this is the moral of the story, the obtaining of the reasons for the ass to eat either bale of hay is not (or not usually anyway) in question. For present purposes, however, this *is* the very question. According to internalism (in its actualist guise), it would appear that A really has *no reason at all* to perform either action as neither one of them is potentially explanatory. I take it that it would be tempting for the internalist (as indeed it would be for anyone) to say that the ass does have a reason to plump for an option in this case. The principle appealed to here might be stated as follows:

- (T)        If A is faced with two incompatible actions,  $\phi$ -ing and  $\psi$ -ing, and the strength of the reason to  $\phi$  is exactly equal to the strength of the reason to  $\psi$ , and A has no stronger reason to do anything else, then A has (good) reason to plump for either action.

This seems to be true even for someone who has no disposition whatsoever to plump; and who, on the current internalist proposal, has no reason to plump. (T), it seems, is true of everyone, regardless of what they are or would be motivated to do. In other words, in so far as an agent has something in his *S* - *regardless of what those elements are* - he has good reason to plump for a particular option in situations where he is equally most motivated to perform two incompatible actions.

So if we think that this agent has a reason to plump, then in so far as the agent is not actually motivated to plump (perhaps because he believes that the decision is so important that it would show a lack of appreciation for the seriousness of the case to let a simple coin toss settle the matter), we must think that there is something wrong with actualism. It is of course true that some motive of A's would be satisfied by his tossing a coin and subsequently actually doing that which the coin decides he should do, and in this sense possibilism seems to get it right.

(ii) *Weakness of will*

Suppose that after having deliberated soundly on all relevant facts, A concludes that, all things considered, he has most reason to  $\phi$ . By  $\phi$ -ing, A would come to realise a long term goal he has been pursuing for a very long time; he would please all his friends and

family by  $\phi$ -ing, which is something he considers highly valuable; and he would also achieve several short term goals by  $\phi$ -ing - you get the idea. However, A is weak willed. Although he believes that he has most reason to  $\phi$ , he is so tempted to go the pub (which is incompatible with  $\phi$ -ing) that he cannot resist doing so. To be sure, A is irrational. Although there can be little doubt that some of A's desires would be satisfied by going to the pub, the question we must ask the actualist is whether A has *any* reason whatsoever to  $\phi$ ? If the actualist takes (ITf) seriously, I cannot see how the answer could be anything other than 'no'!

I think it would be unwise for the actualist to suggest here that A's not being motivated to  $\phi$  is evidence of the fact that A did not deliberate soundly in the first place. Were he to subscribe to this view, then it would be very hard to see how he could make sense of Williams' claim that

Unless a claim to the effect that an agent has a reason to  $\phi$  can go beyond what that agent is already motivated to do - that is, go beyond this already being motivated to  $\phi$  - then certainly the term will have too narrow a definition. 'A has a reason to  $\phi$ ' means more than 'A is presently disposed to  $\phi$ '.<sup>44</sup>

This of course brings out again the problem of getting the modality right in (IT) (and its various other formulations). On the one hand, the internalist must conceive of (IT) in such a way that it actually provides some way of ruling out certain actions as not being possible without making this a matter of metaphysical possibility. On the other hand, the internalist must, conceive of 'possible' in such a way that it allows him to say that it is possible that A should  $\phi$  even though he is weak willed. Again, this may not be an easy thing to do (it may not even be possible), but it is something an adequate internalist theory must provide; especially since (IT) is supposed to provide one of the fundamental motivations behind internalism.

This brings me to the problem of accidie. Now the phenomenon of accidie is not new to the philosophical scene, but as far as I am aware it has mainly been presented as a threat to another kind of internalism. This kind of internalism, which I earlier called 'moral judgement/motivation internalism', is a theory about the supposedly necessary connection between *moral judgements* and motivation. Here, however, we are concerned with the implications of accidie for the connection between *reasons* and motivation. Jonathan Dancy characterises accidie as follows:

People who suffer from accidie are those who just don't care for a while about things which would normally seem to



them to be perfectly good reasons for action; this is so whether the reasons are moral reasons or more ordinary ones. Depression can be a cause of accidie. The depressive is not deprived of the relevant beliefs by his depression; they just leave him indifferent. He knows that if he doesn't act now he will lose the opportunity he has been working for for two years, but he can't see that this matters.<sup>45</sup>

So if we allow that accidie is possible, what implications does this have for actualism? Well if A is suffering from accidie, then he is not motivated to do anything. And if this is the case there is *nothing* A (or anyone else - deliberating on the behalf of A) could arrive at through sound deliberation - there is no motivation to deliberate *from*. Hence A has no reason to do anything. As a response to this problem, the actualist might suggest that what A has reason to do is determined (or at least controlled), in part, by A's pre-accidious S. But this merely puts off the problem.

Actualism, as defined above, makes essential reference to what the agent *would, as a matter of fact* be motivated to do after having deliberated. So although A may have some motive which would be served or furthered by his  $\phi$ -ing, the problem is that, *as a matter of fact, A is not motivated* to  $\phi$ . What would the actualist say about a case like this? Does A have a reason to  $\phi$ ? Since,



according to actualism, it is a necessary condition of A's having a reason to  $\phi$  that A *would* (after deliberation) be motivated to  $\phi$  it cannot be the case that A has a reason to  $\phi$ . And, since an appeal to A's pre-accidious S will not do the trick, some other strategy must be employed.

Another way for the actualist to get around this problem would be to amend the formula by adding the proviso that the agent in question not be accidious. A somewhat clumsy formulation of this idea would be:

- (A\*)        A has reason to  $\phi$  if and only if A would, *insofar as A is not accidious*, be motivated to  $\phi$  after having deliberated soundly on, and from, the motivations A already has

Some actualists might think that the necessity of this proviso is precluded by the very idea of 'a sound deliberative route'. This would be a mistake however. As I (or rather, Dancy) described accidie, it is an affliction associated with *motivation*, not deliberation. So the possibility remains: if an agent is accidious, he could, although perfectly capable of engaging in sound deliberation (unless, of course, this ability has also been affected by his accidious affliction), find himself in the position of not being motivated to do that which he would be motivated to do were he

not acciduous. Actualists cannot therefore claim that the inclusion of the proviso is redundant.

But even with the proviso, the problem with this particular actualist strategy is that it employs a counterfactual analysis of what reasons an agent has - and this is, as we have seen, ruled out by (AC). An opponent of actualism could thus claim (with some justification) that the introduction of the proviso (in (A\*)) seems not only arbitrary, but self-defeating *from the actualist point of view*. After all, the inclusion of the proviso just amounts to an idealisation of the agent's S: were A not acciduous A would have reason to.... But if we are in the business of analysing the agent's S counterfactually, why stop at ruling out accidie? Why not follow McDowell's lead and claim that, counterfactually, had the agent had a 'proper' upbringing he would have been a *phronimos*, and that therefore what the agent has reason to do is what the *phronimos* would be motivated to do? Even if the internalist could show that *some* counterfactual analyses are legitimate by internalist standards (which I doubt) I don't think there is a non-arbitrary way for the internalist to specify which counterfactual analyses are legitimate and which ones aren't. On what non-arbitrary grounds can the internalist dismiss accidie without dismissing other seemingly relevant facts about, say, an agent's (perhaps traumatic) experiences as a child?

The actualist may respond to this challenge by pointing out that (A\*) is, for all practical purposes, equivalent to

(A\*\*) A has a reason to  $\phi$  if and only if A would, *insofar as A has any motivations at all*, be motivated to  $\phi$  after having deliberated soundly on, and from, the motivations A already has.

However, if this is so, the actualist can claim to have provided a non-arbitrary, principled reason for why we should accept actualism and reject McDowell's suggestion: the principled reason is that (A\*\*) merely requires the agent to have some motivations (i.e. it merely requires that the agent's motivational system be functioning) while McDowell's proposal requires that the agent have particular motivations - i.e. the motivations he would have had if he had been given a proper upbringing.

I don't think this proposal will help the actualist. It may be correct to say that a person who is suffering from a *terminal* form of accidie really has no reason to do anything because this person has simply ceased to be an agent. But accidie, as Dancy characterises it, is a temporary affliction.<sup>46</sup> The question under considerations is whether an agent who temporarily fails to be motivated by what would normally motivate him still has a reason to do the thing in question. Suppose that at time  $t_1 \dots t_4$  and  $t_6 \dots t_n$  A is such that he

would be strongly motivated to  $\phi$  if the opportunity presented itself.<sup>47</sup> The opportunity to  $\phi$  presents itself at  $t_5$  when A is accidious. Does A have a reason (at  $t_5$ ) to  $\phi$  at  $t_5$ ? It seems that even with (A\*\*) in place, the actualist will have to deny this. Even if we concede the possibility that an agent can be accidious whilst retaining some of his motivations (it is certainly possible to read Dancy's characterisation of accidie in this way), the actualist will not be able to recover the common sense conviction that those considerations that don't motivate him (but which would normally motivate him, were he not accidious) still retain their normative force. To illustrate this, consider the following example.

A is fanatical about football. He lives for the game and he is especially passionate about his local Sunday league team. Unfortunately, he is not a very good footballer and he has never actually been selected to play for the team. It's the last game of the season and the team bus is getting ready to leave for the away game against their deadliest rivals. Disaster strikes: the team's star striker has broken his leg and the team is one man short. In desperation the team captain phones A and tells him to hurry down to the club house (where the bus is just about to leave) and to bring his football boots with him. This is the chance A has been waiting for for two years. Unfortunately, for some reason or other, A is accidious. At the moment, A is only motivated to stay home and watch game shows on TV. A is not weak willed - it is not as if

he recognises (or believes) that the best thing for him to do is to get on the bus, but that he is more motivated to stay home and watch TV - rather, his team captain's pleas just leave him cold. Does A have a reason to get on the team bus?

According to (A\*\*) he does not. His motivational set is functioning alright: he is highly motivated to stay home and watch TV, but he is not in the least motivated to get on the team bus, and therefore he has no reason to get on the bus. But this sounds extraordinary! Suppose A's acciduous affliction lasted only for one minute. Had the team captain called before or after this time, A would have jumped at the chance to play for the team and rushed down to club house to catch the bus. Had the team captain made the call at any time between  $t_1...t_4$  and  $t_6...t_n$  A would have had a reason at any of those times to get on the bus. However, at  $t_5$ , he did not have a reason to get on the bus at  $t_5$ . Why not simply say that for any time between  $t_1$  and  $t_n$ , had A received the call at that time he would have had a reason to get on the bus, it's just that he wouldn't have been motivated to do so at  $t_5$ ?

For those who are not persuaded by these considerations, consider next the following case. Suppose A has been informed that he has been selected to play for his team later this afternoon, and that the bus is leaving at  $t_5$ . Consequently, A is, at time  $t_1...t_4$ , highly motivated to get on the team bus at  $t_5$ . However, at some time before  $t_5$ , A's evil brother, B, slips him a pill which causes him

to be such that at  $t_5$  he has no interest whatsoever in getting on the bus. It seems clear to me that B has harmed A in some fairly obvious (and, I take it, uncontroversial) respect, but it is hard to see how actualists could agree with this. Since, according to actualism, A has (at  $t_5$ ) *no* reason to get on the bus at  $t_5$ , how can it be that B's having prevented A from getting on the bus at  $t_5$  constitutes a harm to A? Actualists can at best say that all that's happened is that A's S has changed – not that it's changed for the worse – even from A's perspective! It may be objected that this argument begs the question against actualists since it presupposes something of the form 'If x is harmful to A, then there is a reason for A to avoid x (because x is harmful)' which actualism denies. But this only serves to highlight yet another respect in which actualism clashes with common sense.

This brings me to another, and final, point: can actualism make sense of accidie being an *affliction* – i.e. something which an agent is *suffering* from? Can actualists show that accidie is something we all have good reason to try to avoid? I suppose the actualist will say that we have reason to avoid accidie if and only if we would be appropriately motivated to avoid accidie. Fair enough. But what could possibly motivate us to avoid accidie if not the idea that accidie is a bad thing *precisely because it prevents us from being motivated to do that which we have (perhaps very good) reason to do*? This consideration may in fact invite an even

stronger challenge to the actualist: can actualists make sense of the phenomenon of accidie at all (let alone that accidie is a bad thing)? How can actualists make sense of claims like 'I know that I have good reason to  $\phi$ , but I am just not motivated (at all) to do it'?

A final response to the problem of accidie would be to simply deny the existence of accidie. For example, one might argue that the purported phenomenon is relevantly similar to that of akrasia; and that as such it can be explained away by employing a broadly Socratic strategy. The potential success of this approach obviously depends not only on the two purported phenomena being sufficiently similar, but also on the plausibility of the Socratic strategy employed to deny the existence of akrasia. I do not think that this is an attractive strategy; there seems to be overwhelming empirical evidence that accidie is a real phenomenon.

## **2.9 Actualism and reasonable regret**

If the actualist is right about A's not having a reason to get on the bus at  $t_5$  because he was accidious at the time, it is hard to see how A could later reasonably regret having done something he had no reason to do at the time.<sup>48</sup> It seems plausible to say that A could reasonably regret not having gotten on the bus (at least in cases where he could have gotten out of his accidious state by sheer will



power). Intuitively, and as a first approximation, the following principle concerning reasonable regret seems roughly correct:

- (R)            A's regretting (at  $t$ ) not having  $\phi$ -ed at some earlier time,  $t_{-n}$ , is reasonable only if there was a reason for A (at  $t_{-n}$ ) to  $\phi$  at  $t_{-n}$ .

Actualists will no doubt reject (R). In what follows I shall not be arguing that (R) is correct, or that it is reasonable for A to regret not having gotten on the bus; rather, I want to focus on what actualists have to say about reasonable regret.

Since actualists believe that what an agent has a reason to do is completely determined by what he is or would be motivated to do, actualists must, I think, hold that the correct principle governing reasons for regret is

- (R\*)           A's regretting (at  $t$ ) not having  $\phi$ -ed at some earlier time,  $t_{-n}$ , is reasonable (if and) only if A is, or would be, motivated (at  $t$ ) to regret not having  $\phi$ -ed at  $t_{-n}$ .

I assume actualists are committed to (R\*) (or something very similar) in virtue of their general account of reasons. My question is whether (R\*) is a plausible account of reasonable regret?



The first thing to note about (R\*) is that there's something odd about the idea of being motivated to regret something. We rarely, if ever, come to be *motivated* to regret something. Regret is a propositional attitude – usually, if not always, directed at some proposition about a *past* action of ours – which we normally *spontaneously* acquire after having judged or accepted that a previous action of ours had some undesirable property or properties. I say 'normally' because there may be cases of non-spontaneous, or 'instrumental' regret. If someone offered A a large sum of money for regretting having  $\phi$ -ed, A might as a result find himself motivated to regret  $\phi$ -ing (even if he never actually  $\phi$ -ed!).<sup>49</sup> However, even if A manages to bring himself to regret having  $\phi$ -ed (through hypnosis or psychotherapy perhaps), the resulting regret (if it deserves to be called that) can hardly be called spontaneous. In the case just described, A is motivated to regret having  $\phi$ -ed because he is motivated to get his hands on the large sum of money and he recognises that a necessary means to get the money is to become motivated to regret having  $\phi$ -ed. Although spontaneous regret is never motivated in this way, this does not mean that spontaneous (non-instrumental) regret cannot be reasonable or unreasonable. If, for instance, I were to spontaneously regret having  $\phi$ -ed although I recognised that I was warranted in believing that  $\phi$ -ing was the best thing I could have done given the circumstances at the time, my regret would be

unreasonable.<sup>50</sup> Or, to take another example, even though I am not in the least *motivated* at present to regret having abandoned my French studies in school, whenever I think about it I *do* regret having done so. I believe my regret is reasonable. On the actualist account, however, an agent can have a reason to  $\phi$  only if  $\phi$  is capable of being the 'object' of A's motivation. Since spontaneous regret can never be motivated – i.e. spontaneous regret can never be the object of someone's motivation – spontaneous regret can never be reasonable according to actualism. This, I believe, is deeply counterintuitive.

The actualist may respond by saying that regret (spontaneous or otherwise) simply *is* a motivational state. A's regretting having  $\phi$ -ed simply *is* A's being motivated *in a particular way* (to, e.g., hide the fact that he  $\phi$ -ed, apologise for having  $\phi$ -ed, ensure that he won't  $\phi$  again etc). But if this is so, actualists can easily accommodate the claim that spontaneous regret can be reasonable. Actualists can claim that if A's spontaneous regret is, or would be, sustained after having deliberated soundly on all relevant facts, A's regret is reasonable.

This cannot be right, however. It is a mistake to equate regret with being motivated. As I said earlier, (spontaneous) regret is a propositional attitude (if not obviously so) whereas motivation is not. When we say 'I regret having  $\phi$ -ed', what we are saying is equivalent to the 'propositional' 'I regret that I  $\phi$ -ed'. There is no

such equivalence for 'I am motivated to  $\phi$ '. Hence regret cannot consist simply in being motivated in a certain way. It is no doubt true that A's regretting having  $\phi$ -ed can *produce* in him the motivation to do certain things, but this is of no use to the actualist since such a motivation is a *consequence* of A's regret.

But maybe this is all the actualists needs. Suppose that (necessarily?) an agent's spontaneously regretting having  $\phi$ -ed entails his being motivated in certain ways. If regret is always (necessarily?) accompanied by motivation, does this not imply that if an agent regrets having  $\phi$ -ed, then he has a reason to regret having  $\phi$ -ed? I don't think that it does. What the agent regrets is one thing, and what he is motivated to do is another. A's regretting having been rude to B may produce in A the motivation to apologise to B, but this cannot, on the actualist account, show that A's regret is reasonable. At best it can show that A has a reason to apologise to B.

In the end, I think we should reject the actualist account of reasonable regret; and since this account seems to be implied by the actualist account of reasons *simpliciter*, I think the preceding argument provides us with yet another reason to reject actualism.

## 2.10 A final argument

I said earlier that Williams says that the supposed truth of the interrelation thesis provides one of the fundamental motivations behind the internalist account. I want to end this chapter by considering what Williams takes to be the second fundamental motivation of the internalist account. Williams says:

There are many things we can say to people who lack appropriate items in their *S*. Suppose for instance, I think someone (I use 'ought' in an unspecific way here) ought to be nicer to his wife. I say, 'You have a reason to be nicer to her'. He says, 'What reason?' I say, 'Because she is your wife.' He says - and he is very hard case - 'I don't care. Don't you understand? I really do not care.' I try various things on him, and try to involve him in this business; and I find that he really is a hard case: there is *nothing* in his motivational set that gives him a reason to be nicer to his wife as things are.<sup>51</sup>

There are many things I can say about or to this man: that he is ungrateful, inconsiderate, hard, sexist, nasty, selfish, brutal, and many other disadvantageous things. I shall presumably say, whatever else I say, that it would be better if he were nicer to her. There is one specific thing the external reasons theorist wants me to

say, that the man has a reason to be nicer. Or, rather, the external reasons theorist *may* want me to say this: one of the mysterious things about the denial of internalism lies precisely in the fact that it leaves it quite obscure when this form of words is thought to be appropriate. But if it is thought to be appropriate, what is supposed to make it appropriate, as opposed to (or in addition to) all those other things that may be said? The question is: what is the difference supposed to be between saying that the agent has a reason to act more considerately, and saying one of the many other things we can say to people whose behaviour does not accord with what we think it should be? As, for instance, that it would be better if they acted otherwise.<sup>52</sup>

An obvious and elegant response to Williams' question is supplied by Parfit. He says:

We might answer: 'The difference is that, if we merely said that it would be better if this man acted more considerately, we would not be claiming that, as we believe and you deny, he has reason to do so.'<sup>53</sup>

I think it will be instructive to pause for a bit and consider in slightly more detail Parfit's response to Williams. According to Parfit, Williams seems to think that in order for a reason statement to be true it must be the case that 'such reasons would have to get *leverage* on people, by motivating them to act differently.'<sup>54</sup> But, Parfit says, this conception of external reasons is 'too utilitarian'. When we believe that other people have reasons for caring, or for acting, we do not have these beliefs as a way of affecting those people. Our aim is, not influence, but truth.<sup>55</sup> Consider Williams' question about what external reasons could do to people who are not responsive to them:

Unless we are given an answer to that question, I, for one, find it hard to resist Nietzsche's plausible interpretation, that the desire of philosophy to find a way in which morality can be guaranteed to get beyond merely *designating* the vile and recalcitrant, to transfixing them or getting inside them, is only a fantasy of *ressentiment*, a magical project to make a wish and its words into a coercive power.<sup>56</sup>

Consider next Parfit's response:

Williams assumes that claims about reasons could achieve only two things. If such claims cannot get inside people, by inducing them to act differently, they can only designate these people. On the first alternative, these claims would have motivating force. On the second, they would be merely classificatory, since their meaning would be only that, if these people were not so vile, or were in some other way different, they *would* act differently. ...however, there is a third possibility. Even when such claims do not have motivating force, they could be more than merely classificatory. They could still have *normative* force. Perhaps these people *should* act differently.<sup>57</sup>

It is easy to appreciate what Parfit is aiming at here although I believe he may have overshot his target; indeed his last comment only obscures the issue. Saying that A had a reason to act differently is one thing; saying that he *should* have acted differently is another. Reason statements (true ones, anyway) have normative force in their own right - that's part and parcel of what it is to be a (true) reason statement and it is unwise to try to highlight this fact by introducing the word 'should'. It can certainly be true that A had a reason to act differently while it is false that he should have acted differently. Although these people may have a reason to act differently, more needs to be said in order to show that the stronger

claim 'they should have acted differently' is true or appropriate. Externalists can of course maintain that moral considerations provide all agents with reasons for action, but this does not mean that these externalists are committed to the view that, necessarily, all agents should act morally. Perhaps the same externalists also believe that considerations of prudence may outweigh moral considerations. So again, perhaps the agent in question did what he should have done, but there was some consideration in virtue of which he had a reason to act differently although this reason was outweighed by other considerations. At this stage we should be careful to keep reasons statements firmly separated from 'should' or 'ought' statements. These issues will be of crucial importance in chapter four. With these considerations in mind, let us return to Williams' challenge.

Why is it appropriate to say to the insensitive husband that he has a reason to treat his wife better? A very compelling answer to this question is provided by Scanlon. He says:

The supposition is that...there is nothing in the man's "subjective motivational set" that would be served by changing his ways. He is, however, the kind of person about whom Williams would allow us to say that he is inconsiderate, cruel, insensitive, and so on. These criticisms do involve accusing him of a kind of deficiency,



namely a failure to be moved by certain considerations that we regard as reasons. (What else is it to be inconsiderate, cruel, insensitive, and so on?) If it is a deficiency for the man to fail to see these considerations as reasons, it would seem that they must be reasons for him. (If they are not, how can it be a deficiency for him to fail to recognize them?) Why not conclude, then, that the man has reason to treat his wife better....<sup>58</sup>

Williams responds to this criticism as follows:

Let [N] stand in for some normative term: if the critic expresses himself by saying "There is a reason for this man to behave differently to these people", then what he says is of the form "there are considerations about these people's welfare, interests, and so on such that it is [N] that this man should treat those considerations as reasons." What can we take [N] to be? It does not seem to me that there is anything in this way of putting the situation which takes us beyond understanding [N] as - very roughly speaking - "better." We can make this significantly more determinate by explaining that the improvement would lie in the agent's coming to count as reasons considerations which we, other citizens, humane

people in general, count as reasons, but while this may help to explain why we, as critics, express ourselves by saying that "There is a reason for A to behave differently", it does not make that statement, or the [N] that it implicitly contains, any more a matter of A's reasons.<sup>59</sup>

I cannot see how this response in any way meets Scanlon's challenge. What is this mysterious normative term [N] supposed to add to the story? For Williams to insist that such a term is implicit in the claim that this person has reason to act differently seems to be nothing but a rather desperate attempt to save his theory. To say of someone that he has reason to behave differently toward these people does not involve the introduction of some new and mysterious normative term. Williams asks 'What can we take [N] to be?' - the correct answer, I believe, is that [N] is nothing other than 'true'! To say of some proposed action that it would be better that it be performed than some alternative, is to say that there is more reason to perform it than the alternative.

This term [N] may or may not be contained (explicitly or implicitly) in the claim we are considering. The natural, and, I believe, correct way of interpreting the claim is simply to say there are considerations about these people's welfare and interests *such that they give this person a reason (or reasons) to treat them*

*differently*. There is no mystery here. This is all that is meant by saying that this person has reason to act differently.

This response also suggests a way to respond to Williams' more general challenge to the externalist:

What is it the agent comes to believe when he comes to believe he has a reason to  $\phi$ ? If he becomes persuaded of this supposedly external truth, so that the reason does then enter his S, what is that he has come to believe? This question presents a challenge to the external reasons theorist.<sup>60</sup>

Here Williams is presenting a challenge to the external reasons theorist in the form of a sceptical question about the *meaning* of external reason statements (as opposed to the truth or falsity of such statements). The challenge consists in the claim 'that the sense of external reason statements is [not] in the least clear....'<sup>61</sup> But why should the externalist take this challenge seriously? Again, there is no mystery here. As John Skorupski puts it, the content of beliefs about external reasons 'seems perfectly clear. It does not cry out for analysis in terms which eliminate the concept of a *reason*.'<sup>62</sup> Of course, it is a substantive issue whether particular external reason statements are true or false, but their meaning seems clear enough. When A comes to believe the external reason

statement that he has a reason to  $\phi$ , he comes to believe nothing more than, simply, that he has a reason to  $\phi$ .

### **2.11 Summary and conclusion**

I think the points discussed in this chapter give us good reason (as it were) to reject internalism. Given the complexity and limitations of our individual psychologies and the vagaries of our actual motivational make up, the internalist model is simply not very attractive. Externalists, it seems to me, do better in every respect. They can quite happily admit that A can have a reason to  $\phi$  at the same time as he has a reason to not- $\phi$ ; if A is weak willed, accidious, or indecisive, then that's one thing - what he has a reason to do is another. The psychological or motivational state of a particular agent has very little, if anything, to do with what he has a reason to do (although, as Scanlon puts it, external reasons may have 'subjective conditions'). For externalists, the truth of a particular reasons statement does not (in general) entail (and nor is it entailed by) the agent's having a desire that would served by his  $\phi$ -ing.

<sup>1</sup>Reprinted in Williams (1981).

<sup>2</sup>Williams (1995b), p. 186.

<sup>3</sup>McDowell (1995), p. 68. This passage obscures an important issue. McDowell's use of the word 'ethical' is, in this context, somewhat misleading. Williams has never, to my knowledge anyway, doubted the existence of *ethical* reasons. What he is suspicious of are *moral* reasons. It is after all *morality* that is the 'peculiar institution' (see Williams (1985), Ch. 10). For Williams, moral considerations (or, more accurately, moral obligations) form a subset of ethical considerations and their peculiarity consists in their apparently built-in claim to categoricity, to a supposed ability to outweigh or perhaps even silence all other ethical considerations.

<sup>4</sup>It could be argued that this is actually not so. It may be the case that those who are simply unable (whatever that turns out to mean) to see moral considerations as being reason-giving, simply fall outside the scope of morality *because of, or in virtue of, their very inability* to see moral considerations as reason-giving. There is something to be said for this view, but I cannot take it into consideration in this dissertation. For an important discussion of this topic, see Skorupski (forthcoming).

<sup>5</sup>Unless of course they have some derivative, or 'proleptic' (as Williams puts it) reason to refrain from doing such things. The sadistic amoralist would have a reason not to torture some potential victim if his doing so would lead to something he is averse to (like the withdrawal of respect from people whose respect he desires for example). See 'Internal Reasons and the Obscurity of Blame' in Williams (1995).

<sup>6</sup>See Dancy (2000), p. 31.

<sup>7</sup>See Darwall (1997).

<sup>8</sup>In Darwall's taxonomy, the necessity operators have a narrow scope, i.e. they govern only the consequents of the conditionals. I think this is an oversight on Darwall's part. These principles must be taken to have wide scope necessity operators, i.e. the necessity operators must be taken to govern the whole conditional.

<sup>9</sup>'Internalism and the Obscurity of Blame' in Williams (1995); and 'Replies' in Altham and Harrison (eds); and 'Postscript: Some Further Notes on Internal and External Reasons' in Millgram (ed.)

<sup>10</sup>Williams (1981), p. 102.

<sup>11</sup>*Ibid.*, p. 105. It is far from clear whether Williams thinks that all elements in an agent's S must be, in some sense or other, non-cognitive. Suppose cognitivism about value is correct, where does that leave an agent's 'disposition of evaluation'? If there are truths about values, in the sense that we can be wrong about what is valuable, then someone's being disposed to value things (actions,

persons, ideals etc.) in a certain way seems be redundant with respect to determining what he has reason to do. I shall return to this point in the next chapter.

<sup>12</sup>*Ibid.*

<sup>13</sup>*Ibid.*, p. 104.

<sup>14</sup>Williams (1995), p. 38.

<sup>15</sup>*Ibid.*, p. 37.

<sup>16</sup>Williams (1981), p. 102.

<sup>17</sup>Williams (1995), pp. 38-9.

<sup>18</sup>*Ibid.*, p. 38.

<sup>19</sup>Williams (1981), pp 102-4.

<sup>20</sup>See Williams (1995), p. 35.

<sup>21</sup>*Ibid.* My italics.

<sup>22</sup>*Ibid.*, p. 36. My italics.

<sup>23</sup>*Ibid.*, p. 40. My italics.

<sup>24</sup>I have benefited greatly from Parfit's discussion of this matter. See Parfit (1997). In that paper Parfit claims that Williams, in the end, actually rejects what Parfit calls 'Analytical Internalism' - see p. 110n21.

<sup>25</sup>Even if Williams is not trying to analyse the *meaning* of reasons statements, we still need to know what it means to say of someone that he or she 'could reach the conclusion to  $\phi$ '.

<sup>26</sup>Williams (1995), p. 35.

<sup>27</sup>Williams (1981), p. 103.

<sup>28</sup>Possibilism is essentially a slightly expanded version of what Williams calls 'the sub-Humean model' of practical reason. See Williams (1981), pp. 101-2.

<sup>29</sup>See Williams (1981), p. 107.

<sup>30</sup>*Ibid.*, p. 105.

<sup>31</sup>See Smith, p. 105.

<sup>32</sup>*Ibid.*

<sup>33</sup>*Ibid.*, p. 114.

<sup>34</sup>'Should' should here be understood non-normatively.

<sup>35</sup>Williams (1995), p. 38.

<sup>36</sup>Williams (1981), p. 107.

<sup>37</sup>*Ibid.*, 103.

<sup>38</sup>The problematic implication of this apparent truth has recently been highlighted by Jonathan Dancy (2000). In so far as A  $\phi$ 's intentionally, a satisfactory explanation of his doing so must also, as Williams points out, include the fact that he accepts (or believes) he has a reason to  $\phi$ ; how else could his  $\phi$ -ing be intentional? But if it is necessary to include A's acceptance of the fact that he has reason to  $\phi$  in the explanation of his  $\phi$ -ing, then it is far from clear what role the fact that A has a reason to  $\phi$  is supposed to play in the explanation of A's  $\phi$ -ing. Does it even play any role in the explanation of his action? It seems that what does the explaining is



A's acceptance of a certain proposition rather than the truth of that proposition. Consider the case where A  $\phi$ 's because he mistakenly believes he has reason to  $\phi$ . What is the proper explanation of A's  $\phi$ -ing in this case? Well, whatever the correct explanation is, it cannot include the fact that A had a reason to  $\phi$ , because *ex hypothesi* there was no such fact. Rather, the correct explanation of A's  $\phi$ -ing seems to be that he (mistakenly) believed he has reason to  $\phi$ . But if it is the fact that A (correctly or incorrectly) believed himself to have a reason to do something that explains his doing it, then, as Dancy rightly points out, it is hard to see how somebody could ever be said to act *for a good reason* rather than because he believed there to be good reason to do it.

<sup>39</sup>Williams (1981), p. 107.

<sup>40</sup>I owe this point to Hugh Mellor.

<sup>41</sup>See McDowell (1995).

<sup>42</sup>Williams (1995b), p. 190. It is not exactly clear what Williams' target is here. On the one hand Williams can be read as saying that whether or not the phronimos would  $\phi$  in circumstance C is irrelevant to whether A has a reason to  $\phi$  in circumstance C. The passage could also be read as saying that if the phronimos would  $\phi$  in circumstance C, then A has *some* reason to  $\phi$  in circumstance C, but, given that A relevantly fails to be a phronimos, he also has *some* reason not to  $\phi$  in circumstance C. I suspect that the former interpretation is the correct one.

<sup>43</sup>From now on, I shall not always make a distinction between (IT) and (IT\*) but use them - and their amended versions - interchangeably depending on context.

<sup>44</sup>Williams (1995), p. 36.

<sup>45</sup>Dancy (1993), p. 5.

<sup>46</sup>We can see how questions about vagueness might creep in here, but I shall not address this concern.

<sup>47</sup>In what follows, whenever I speak about motivation, or an agent's being motivated, I shall take 'after having deliberated soundly on, and from, his existing motivations' as read.

<sup>48</sup>'Reasonable' is quite an elusive term. According to some authors, it is reasonable for A to  $\phi$  if and only if  $\phi$ -ing is what A has *most* reason to do. Others think that it is reasonable for A to  $\phi$  if and only if A *believes* he has most reason to  $\phi$ . Others still have argued that it is reasonable for A to  $\phi$  if and only if A is *warranted* in believing that he has most reason to  $\phi$ . To this extent, 'reasonable' shares the same features as 'rational'. By 'reasonable' I mean nothing more than 'is supported by a reason'. Thus, 'A's regret is reasonable' should be read as 'A's regret is supported by a reason'. This is mere stipulation on my part and it is done merely for the sake of ease of exposition.

<sup>49</sup>If this is correct, (R) must be false. I should also note that some philosophers would deny that A really has a reason to regret having  $\phi$ -ed in the case I have just described (even though he is highly motivated to do so). These philosophers think that A's being (appropriately) motivated to regret having  $\phi$ -ed gives him a reason to try to get himself to regret having  $\phi$ -ed. This, they argue, is different from having a reason to regret having  $\phi$ -ed. I shall not pursue this line of thinking.

<sup>50</sup>There is a sense in which my regretting having acted in particular way could still be reasonable even though I am warranted in believing that what I did was the best thing I could have done at the time. My regret could be directed 'at the world' so to speak, and not at my own action. Roughly, my regret in such a case should be seen as an expression of my belief that it was *unfortunate* that I had to do what I did. This is not a criticism of my own action; rather it is a criticism of the circumstances in which my act was performed: 'I regret having slapped him, but he was hysterical...'

<sup>51</sup>I take it that Williams intends to say 'there is nothing in his motivational set *which would be served or furthered* by his being nicer to his wife.' To say that there's nothing in his S *that gives him a reason* to be nicer to his wife seems out of place here. If that remark is supposed to mean that there simply is no reason for him to treat his wife better, then this would obviously be question begging against the externalist.

<sup>52</sup>Williams (1995), pp. 39-40.

<sup>53</sup>Parfit (1997), p. 109.

<sup>54</sup>*Ibid.*, p. 111.

<sup>55</sup>*Ibid.*

<sup>56</sup>Williams (1995b), p. 216.

<sup>57</sup>Parfit (1997), p. 112.

<sup>58</sup>Scanlon (1998), pp. 366-7.

<sup>59</sup>Williams in Millgram (ed.), p. 96.

<sup>60</sup>Williams (1995), p. 39.

<sup>61</sup>*Ibid.*, p. 40.

<sup>62</sup>Skorupski (forthcoming).



## CHAPTER THREE: BELIEFS, DESIRES, AND MOTIVATION

### 3.1 Introduction

In the previous chapter I presented some arguments which I believe seriously undermine the plausibility of internalism. Some of these arguments were designed to show that simply having a motive that consists in, or which would be served by,  $\phi$ -ing is not sufficient to provide an agent with a reason to  $\phi$  - even when that motive is not the product of false belief. Other arguments were constructed to show that A's being motivated to  $\phi$  is not a necessary condition for its being the case that A has a reason to  $\phi$ . The internalist can of course maintain that the core idea of internalism has not yet been shown to be false. In a recent comment on the internalism/externalism debate, Williams says:

The formulation of the internalist position which I now favour is: A has a reason to  $\Phi$  only if there is a *sound deliberative route* from A's subjective motivational set...to A's  $\Phi$ -ing. Whether this is also a sufficient condition of A's having a reason to  $\Phi$  is a question which I have left aside; the essence of the internalist position is that it is a necessary condition.<sup>1</sup>

The internalist might thus be content with the idea that internalism is merely a negative doctrine - that internalism should be thought of merely as specifying a necessary condition for the obtaining of a reason (for any agent). To this extent, then, since we have already dismissed the idea that being motivated to  $\phi$  provides a necessary condition for the obtaining of a reason to  $\phi$ , if the internalist still insists that he is in the business of specifying the necessary conditions for the obtaining of reasons, it seems he is committed to the idea that the necessary condition consists not in the agent's being motivated to  $\phi$ , but in the agent's having some motive which would be satisfied by his  $\phi$ -ing.

I have two aims in this chapter. First, I want to investigate whether internalists can allow the idea that beliefs can, in themselves, be elements (i.e. motives) of an agent's motivational set. I shall then proceed to show why I think this is not the case. Second, since I believe internalists are ultimately committed to the thesis that an agent's motives simply are his desires (or desire-like states), I shall try to put some pressure on the idea that desires are reason-giving, and suggest that desires can themselves be reasonable or unreasonable. Let us begin by looking at the role of beliefs in internalism.

### 3.2 The role of beliefs in internalism

Sometimes Williams writes as if he thinks that beliefs can motivate in their own right; and if they can do this, there seems to be no *prima facie* reason for denying that beliefs are 'motives' - i.e. that they can be elements in an agent's S. But how should we understand 'motive' here? Here I shall follow John Skorupski:

Let us say that a motive is whatever could be adduced, in our everyday explanations of intentional action, as explaining why the action was done. The question 'What was A's motive in doing this?' and 'What was A's reason for doing this?' are effectively the same question.<sup>2</sup>

Slightly reformulated, then, the question is: can beliefs be motives? Again, Williams seems to think that they can be. The problem for Williams is that he cannot accommodate this intuitively plausible thought into his version of internalism; or so I shall argue anyway. I hope to show that if Williams wants to hold on to the idea that beliefs can generate reasons, he will be forced to reject, or at least substantially revise, some of the more important claims that underwrite his internalism. We should also remember the motive/motivation distinction. It seems clear enough that Williams believes that beliefs can motivate, and I think we should grant him that this is true. However, if beliefs can motivate, does that also

make them (i.e. the beliefs, or, more accurately, their content) motives?

Williams writes:

Does believing that a particular consideration is a reason to act in a particular way provide, or indeed constitute, a motivation to act? ... Let us grant that it does - this claim indeed seems plausible, so long at least as the connexion between such beliefs and the disposition to act is not tightened to that unnecessary degree which excludes *akrasia*. The claim is in fact so plausible, that this agent, with this belief, appears to be one about whom, now, an *internal* reason statement could truly be made: he is one with an appropriate motivation in his *S*.<sup>3</sup> (Call this passage  $\alpha$ )

I think the inclusion of the *akrasia* proviso is misplaced here, but I shall defer my discussion of why I think so until later. For present purposes however, let us ignore this point and apply this passage to the story about Owen Wingrave.

In the original story by Henry James, Owen has no desire whatsoever to join the army despite his family's insistence on the fact that family tradition requires him to do so. In 'Internal and External Reasons' Williams uses the case of Owen Wingrave to

illustrate the difference between internalism and externalism.

When Owen's family insist that he has a reason to join the army despite the fact that Owen has no desire whatsoever to join the army (which of course they are fully aware of), they take an externalist view of reasons. Suppose we change the story so that Owen has now come to *believe* that considerations of family tradition and honour provide him with a reason to join the army. If we take Williams' suggestions in  $\alpha$  seriously, it would now appear to be true of Owen that he does have reason to join the army. But in virtue of *what* is it now true that he has reason to join the army? - in virtue of what is Owen's motivation (now) 'appropriate'?

Consider the following argument<sup>4</sup>:

- (1) Owen (O) believes that [the fact that family tradition bids O to join the army gives O a reason to join the army]. (Call the contents of this belief 'p') (Assumption)
- (2) O's belief that p provides him with a motivation to join the army. (From  $\alpha$ )
- (3) If O is motivated to join the army then O has a reason to join the army. (From  $\alpha$ )

From (1), (2), and (3) Williams may want to conclude

(4) O has reason to join the army.

I find this argument puzzling. There seem to be a number of ways in which it is faulty. The first thing we should note is that despite initial appearances, (2) is not very plausible - not if we read it as a conceptual truth about the connection between believing oneself to have a reason to do something and being motivated to do that thing. It certainly seems possible that an agent can believe (and, indeed, know) that he has some motive which consists in, or which would be served by, his  $\phi$ -ing without being motivated to  $\phi$  even though he is presented with an opportunity to  $\phi$ . Even more importantly, this lack of motivation can persist without there being anything wrong with the agent; i.e. the lack of motivation does not have to be explained by the agent's being accidius or akratic.

Consider the following case. Suppose A believes that he has a motive that would be served by his  $\phi$ -ing and that he has a motive that would be served by his  $\psi$ -ing. It seems possible, and in certain circumstances very likely, that A would not be motivated to, say,  $\psi$  even though he believes he has a reason to  $\psi$ . Suppose A's  $\phi$ -ing consists in his cashing in his winning lottery ticket (thereby winning a million pounds), and that  $\psi$ -ing consists in his going jogging (thereby promoting his health). Does it follow that if A were aware of this he would be motivated to return his winning lottery ticket as well as to go jogging? I cannot see that it does. I believe it is

perfectly possible that although A believes (and indeed *knows*) that he has a reason to return his lottery ticket and that he has a reason to go jogging he is not at all motivated to do the latter. And, as I said, A can lack the motivation to go jogging without thereby arousing the suspicion that he is suffering from some motivational disorder.<sup>5</sup>

So if we interpret (2) as a supposedly conceptual truth about the connection between normative beliefs and motivation, then since (2) (thus interpreted) is false, the argument is unsound. To give Williams the benefit of the doubt however, we can allow that Owen's belief *does* motivate him, while retaining the conviction that it is a contingent matter whether or not an agent is actually motivated by his or her normative beliefs. But this brings out an important question: how important is the fact that Owen's belief actually motivates him? Would Owen still have a reason to join the army if his belief that he had reason to do so left him cold? I shall return to this question later.

So although I believe that (2) is pretty weak, the weakest link in this argument is (3). In virtue of what is (3) true? Williams seems to suggest that the reason why we can derive (4) is that Owen's motivation is (now), in some sense or other, *appropriate*. What does he mean by this? How, or in virtue of what, is Owen's motivation appropriate? We can agree with Williams that the internalist should hold the following to be true:



- (3.5) If O's motivation to join the army is of the appropriate kind then O has a reason to join the army.

I take it that this is what it means (on the internalist view) to say of someone that he has 'an *appropriate* motivation in his S'. However, it is far from clear whether this move actually adds anything to the story. What does it mean, on Williams' account, to say of someone that she is appropriately motivated? Well, presumably it means that this person has a reason to do the thing in question. What else *could* it mean? To say of someone that she is appropriately motivated to do something is to say that she has a reason to do that thing. However, Owen's being motivated to  $\phi$  and Owen's being *appropriately* motivated to  $\phi$  are not the same thing. So how does Williams get to the conclusion that Owen has reason to  $\phi$ ? What Williams needs to show is *why* it is that Owen's motivation is appropriate (i.e. why he has a reason to join the army) - not just assert it.

More importantly, (3) is false if the belief that gave rise to Owen's motivation is false. Remember Williams' claim that if an agent's motivation to  $\phi$  is the result of his holding a false belief, then that motivation cannot make it the case that he has a reason to  $\phi$ . Let us put this as

- (5) If O's motivation to join the army is the product of a false belief, then (*ceteris paribus*) it is not the case that O has a reason to join the army.

The inclusion of the *ceteris paribus* clause is intended to serve merely as a reminder that we are (at this stage, anyway) presupposing that Owen has no other motive besides his belief that p to join the army. So the antecedent in (3) is not strong enough to guarantee the truth of the consequent. What (3) needs, if it is to guarantee the truth of its consequent is that Owen's belief that p is *true*. But no such strengthening of the antecedent has been offered. (1) simply says that Owen believes that p - it says nothing about whether he truly or falsely believes that p. Now, since, according to (2), Owen's motivation to join the army was provided by his *belief* that p, and according to (4), Owen does have a reason to join the army, we can now (via (5)) derive

- (6) It is not the case that O's belief that p (which motivated him to join the army) is false. (From 2, 4, and 5)

But from (6) it must surely follow that

- (7) O's belief that p is true.

And from (7) it must equally surely follow that

(8) p is true.

In other words, it is true that the fact that family tradition bids Owen to join the army gives Owen a reason to join the army. The crucial question now becomes: in virtue of what is p true? As far as I can see there are only two possibilities. Either

(9) p is true independently of any elements in O's motivational set.

or

(10) p is true in virtue of some element(s) in O's motivational set

Clearly Williams cannot accept (9). According to (9), the proposition 'the fact that family tradition bids O to join the army gives O a reason to join the army' is true *independently of any elements in O's motivational set*. But this, of course, is nothing short of externalism! So Williams must reject (9), and accept (10).

The next question we need to consider is whether the element in O's motivational set referred to in (10) is a belief or a desire.<sup>6</sup>

Thus we have

(11) The element in O's motivational set which makes p true is a belief

or

(12) The element in O's motivational set which makes p true is a desire

Let us investigate the merits of (11) first. The first question we need to ask with respect to (11) is *which* belief it is that makes p true. Again, there are two possibilities:

(13) It is O's believing p that makes p true

or

(14) It is some belief of O's, other than O's belief that p, that makes p true.

If (13) is true, then beliefs about what we have reason to do are either infallible, self-verifying, or in some other way 'extension-determining'. Since our having these beliefs is what makes them true, we could never have false beliefs about what we have reason to do (at least in cases where all our relevant non-normative beliefs are correct). Thus it seems beliefs about what we have reason to do are of the same kind as beliefs about our own sensory states: if I believe that x appears red to me, then x appears red to me; if I believe that I am in pain then I am in pain etc.. However, we

should be careful to point out that there is an important disanalogy between the case we are currently discussing and the case of the mistaken belief about the glass containing gin (which we encountered in chapter two). In the gin case, the agent has no belief about *what he has reason to do*; he simply has a false belief about the contents of the glass before him. In that case, his desire (or motivation) to drink the stuff in the glass before him *does* result from a false belief. As a result he has no reason to drink the stuff before him. In the present case however, the content of this agent's belief *is a proposition about what reasons he has* - and this is the crucial difference.

However, Williams explicitly (and rightly) rejects the idea that our normative beliefs are self-verifying: '[O] may falsely believe an internal reason statement about himself...' <sup>17</sup> Let us add this proposition to the argument:

- (15) O may falsely believe an internal reason statement about himself.

This rules out (13): if we can falsely believe internal reasons statements about ourselves, then it cannot simply be *our believing* ourselves to have a reason that makes it the case that we actually do have a reason to do the thing in question. Then, in so far as Williams wants to hold on to (11), he will have to accept (14): it is

some belief of O's, other than O's belief that p, that makes p true.

The question now becomes: what could this belief possibly be?

Remember what p stands for: the fact that family tradition bids O to join the army gives O a reason to join the army. What other belief of O's could make this belief true? One strategy would be to say that there is some other, more 'fundamental' belief of O's that makes p true. For example, the belief that makes p true is O's belief that O has reason to do whatever family tradition bids him to do (call the content of this belief 'q'). But this merely puts off the problem. If q is false, then, assuming as before that O has no other reason to join the army, O has no reason to join the army. So if O's belief that p is true in virtue of O's believing q, then q must be true. So what makes *this* belief true? Obviously an infinite regress threatens here.

As I see it, there are only two ways for Williams to get round this problem. Either he holds that q, or some other, even more fundamental normative belief of Owen's is *basic*; or, alternatively, he abandons (11) in favour of (12). If he chooses the former, he will have to accept that (15) is true only of non-basic (i.e. derivative) normative beliefs. Basic normative beliefs (whatever they turn out to be), on the other hand, are infallible. Let us add this as a premise.

(15\*) O cannot falsely believe a *basic* internal reason statement about himself.

To investigate the merits of this view, let us introduce the idea of a basic normative belief and a corresponding basic normative reason and distinguish this kind of belief/reason pair from instrumental beliefs/reasons pairs by adding the following to our stock of premises:

(16) If O's belief that he has a reason to  $\phi$  is a basic normative belief, then O has a (basic) reason to  $\phi$ .

### 3.3 The content of basic normative beliefs

So far I have not said much about the content of these allegedly basic normative beliefs. Recall that, according to Williams, if a person believes that he has a reason to do something, then there must be some determinate consideration in virtue of which he takes himself to have that reason. In Owen's case, he believes not only that he has a reason to join the army, but he also believes that the fact that family tradition bids him to join the army is a (or *the*) consideration which provides him with a reason to join the army. We must now ask whether this model also applies to basic normative beliefs. Is there some determinate consideration in



virtue of which Owen believes that he has a reason to do whatever family tradition bids him to do? Suppose Owen believes that it is simply important that family traditions be followed and that this fact gives him a reason to do whatever his family tradition bids him to do. If this is what Owen believes and his belief is genuinely basic, then this belief cannot be the product of some other normative belief he has.

Now, given our characterisation of basic normative beliefs, if Owen's belief that *q* is basic then *q* is true. Importantly however, since Owen's belief that *q* is basic (unlike Owen's belief that *p* which is made true by something else - namely Owen's believing that *q*), *q* must be *made true* by Owen's believing it. A brief recap of the story so far is in order here. The dialectic up to this point have taken us from

(11) The element in O's motivational set which makes *p* true is a belief

to

(14) It is some belief of O's, other than O's belief that *p*, which makes *p* true.

We are now supposing that Owen's belief that *p* is made true by his belief that *q*. If Owen's belief that *q* is true and this belief is not

made true by any other belief of Owen's, then  $q$  must be made true by Owen's believing  $q$ . Thus we arrive at

(16\*) If  $O$ 's belief that  $q$  is basic, then  $q$  is made true by  $O$ 's believing it.

The next question we need to consider is what could and could not be the content of basic normative beliefs. Can one have basic normative beliefs about what other people have reason to do? Suppose Owen believes that since it is important that family traditions be upheld *everyone* has a reason to do that which his or her family tradition bids him or her to do. Suppose that in Owen's family there is a tradition of the oldest son joining the army and for the second oldest to join the clergy. Since Owen is the oldest he believes of his younger brother, Bob, who has no interest whatsoever in doing what family tradition bids him to do, that he has a reason to join the clergy. If this belief of Owen's is true, then Bob has a reason to join the clergy. According to the standard internalist picture, however, 'Bob has a reason to join the clergy', is false since, *ex hypothesi*, there is no motive in Bob's  $S$  which would be satisfied or promoted by his joining the clergy. So it is simply false that the fact that family tradition bids Bob to join the clergy gives Bob a reason to do so. But if this is false, then Owen's belief that there is reason for Bob to join the clergy must also be false.

This in turn shows that Owen's belief that everyone has a reason to do that which their respective family traditions bids them to do is false. And hence this belief of Owen's cannot be basic.

I think this argument strongly suggests that internalists must hold that basic normative beliefs, if there are any, can at most be *de se* beliefs. If by some extraordinary coincidence it so happens that all agents share Owen's conviction that considerations of family tradition are reason giving, then Owen's belief would be true. But it wouldn't be *made true* by Owen's believing it; rather it would be made true by the contingent fact that everyone agrees that upholding family traditions is important. So if basic beliefs are made true by being believed, it would appear that basic beliefs can at best only have the believer as their object.

With the *de se* qualification in mind, let us now return to (16\*). If we, as internal reasons theorists, were to suggest to Owen that the content of his belief is made true by his having that belief, I think it is fairly safe to say that Owen would reply that we have misunderstood him (or his belief). I think he would be likely to say to us: 'You haven't understood what I believe. I believe that *it is the importance of keeping with family tradition* that gives me a reason to do that which family tradition bids me to do - not that I believe it.' In other words, Owen is expressing an externalist thought. Again, we could easily imagine him saying 'What I believe is that the reason I have to honour family tradition obtains

*independently* of my believing it!' How should the internalist respond to this? There seems to be only two possibilities here. Either Owen's belief is false, and thus not basic; or the belief is true, and, consequently, there are external reasons.

Obviously, the internalist must say that Owen's belief is false, thereby denying that Owen's belief is basic. But this seems unsatisfactory *from an internalist point of view*. Here is a case where Owen, in Williams' words, has a 'pattern of emotional reaction' such that he is always motivated only by what he takes to be *bona fide* external reasons; or that Owen's 'disposition of evaluation' is such that when he believes he has a reason to do something, he believes this, at least in part, *because* he believes this reason obtains independently of his being motivated by it; or that Owen has a 'commitment' to what we might call 'the externalist cause'. So even though all these things are true about Owen, internalists are forced, I think, to say that it is false that Owen has a reason to join the army.

It seems reasonable to think that internalists will have to say that most, if not all, basic beliefs are false since it seems equally reasonable to believe that the point just made can be generalised: most (if not all) basic normative beliefs will have the same structure as Owen's belief that *q*; i.e. they all have, as part of their content, the idea that the reasons in question obtain independently of

whether we believe them to - to think otherwise would be to get things back to front (at least for a very large class of such beliefs).

I think this is a powerful argument against the possibility of internalists having recourse to the existence of basic normative beliefs in order to maintain that beliefs can be motives which generate reasons. However, as we shall see, there are still a few arguments available to the internalist who wants to defend the idea that beliefs can be motives. Perhaps it is possible to maintain that an agent can have a non-desire-based basic normative reason to  $\phi$  without *occurrently* believing this to be so. If we allow that there are such things as *tacit* beliefs, then it might be possible for a basic normative reason statement to be true of an agent in virtue of his having a tacit basic normative belief.

### **3.4 Basic normative beliefs as tacit beliefs**

The idea of tacit beliefs raises some very intricate and difficult issues about the epistemology and metaphysics of beliefs and these are issues which I cannot go into detail about here. As a result, I must limit this discussion to what I take to be directly relevant to the issue under consideration. What does it mean to say of someone that he has a tacit basic normative belief? There are some philosophers who doubt the very existence of tacit beliefs,<sup>8</sup> while others think that, at best, we are merely warranted in

*ascribing* tacit beliefs to others (if not to ourselves).<sup>9</sup> However, for those who believe there are such things, the first order of business must surely be to say what they are. A couple of distinctions are in order here. In the literature, philosophers have distinguished between different kinds of beliefs: explicit and implicit; conscious and unconscious; occurrent and non-occurrent, and so on. So where, if at all, do tacit beliefs fit in? Well, the answer to that question, I believe, is largely terminological. The main idea is that tacit beliefs, if there are any, are such that they have never been 'entertained'. Let us call a belief that is currently 'before the mind's eye' - i.e. one that is presently being entertained - an occurrent belief. A non-occurrent belief, then, is a belief that is not presently entertained. However, there are at least two kinds of non-occurrent beliefs. On the one hand there are beliefs which are not currently being entertained but which have, at some point, been entertained and perhaps stored in either long-term or short-term memory. These are the kinds of beliefs we are all familiar with: that  $2 + 2 = 4$ ; that 'b' immediately follows 'a' in the English alphabet; that Monday follows Sunday etc. These beliefs, although non-occurrent must be distinguished from other kinds of non-occurrent beliefs an agent, A, might hold: that A is less than eight feet tall; that  $10,329 > 10,328$ ; and that snow in Stockholm does not instantaneously turn bright orange when it hits the ground.<sup>10</sup> These beliefs have never, we may suppose, been entertained by A. Nonetheless, were

we to ask A whether he believes that he is less than eight feet tall etc. he would immediately reply 'yes' even though he had never entertained that particular belief before. These kinds of beliefs are the ones we might call 'tacit'.

According to a very tentative analysis, A's tacit belief that q consists in A's disposition to assent to q if the occasion arises. However, consider the following case (due to Lycan):

*The opinionated man.* He is a Peircean, in that he abhors being agnostic on any subject, but not enough of a Peircean, in that in him the 'irritation of doubt' triggers not inquiry but a snap judgement. At least, on many occasions when he entertains a proposition for the first time, he immediately affirms the proposition or denies it, depending on what else is going on in his global psychology at the time. (Let us take 'global psychology' broadly here, to include any mental or neurophysiological condition that has psychological influence.) Thus at a time *t* our subject has countless dispositions to judge - determined by his global psychology - but we would not count these as antecedently existing beliefs, however tacit.<sup>11</sup>

So the tentative analysis is inadequate. What we need is some way of getting around the arbitrariness of snap judgements; a way of



showing how A's disposition to judge that  $q$  is not merely capricious:

An obvious move is to posit what Dennett (1975) calls an 'extrapolator-deducer', i.e. a device which operates on [previously entertained] stored formulas, or 'core beliefs', and generates relatively obvious consequences of those [previously entertained] beliefs when the occasion arises. Thus we might suggest that to believe tacitly that  $P$  is to be disposed to judge that  $P$  in virtue of the operation of one's extrapolator-deducer in drawing inferences from pre-existing beliefs.<sup>12</sup>

But this will not do for our basic normative beliefs. Recall that the kind of beliefs we are interested in are supposed to be *basic*, and if we are to respect this basicness we cannot allow that they be inferred from other, even more basic beliefs along the extrapolator-deducer model. In Owen's case we considered the possibility that his belief that  $p$  was made true by his belief that  $q$ , which in turn was made true by being believed. The idea is that the 'making true' relation that holds between Owen's beliefs is modelled on a syllogism of the following kind:

Basic belief: I (Owen) have a reason to do whatever family tradition bids me to do.

Empirical belief: Family tradition bids me to join the army.

Derivative belief: (The fact that family tradition bids me to join the army makes it the case that) I have a reason to join the army.<sup>13</sup>

According to this model, the major premise (the basic belief) in the syllogism is made true by being believed; the minor premise (the empirical belief) is made true by the obtaining of a certain state of affairs which obtains independently of Owen's believing it; and the conclusion (the derivative belief) is made true by the truth of the major and minor premises. If we were to view the extrapolator-deducer model of tacit beliefs as providing a way of securing the existence of basic normative beliefs, we would end up with the wrong structure of the 'making true' relation - the structure would be the reverse of the syllogism just described. Since tacit beliefs are those 'relatively obvious consequences' of some 'core' beliefs, these tacit beliefs, if true, cannot make the 'core' beliefs true. To suppose that they could would be to get things back to front.

There are of course other models of tacit beliefs. According to Field<sup>14</sup>, tacit beliefs are the obvious consequences of one's occurrent or non-occurrent-but-have-been-entertained beliefs. The idea here is that what is to count as the obvious consequences of a particular

belief will be determined by the person *ascribing* the belief to the agent in question. Now one could certainly question whether this is a genuine attempt at specifying a belief of an agent as opposed to merely ascribing it to the agent. Or again, as Lycan points out, one could read Fields proposal as a *deflationary* analysis of tacit belief.<sup>15</sup> But regardless of which interpretation we find more plausible, this model is of no use to us. For our purposes, it is simply a different version of the idea expressed by Dennett. If a basic normative belief is a tacit belief, it had better be a tacit belief of such a kind that it is not merely a *consequence* of some other belief, ascribed or otherwise. I think we should conclude that if there are basic normative beliefs, then we should not think of these as tacit beliefs. There may of course be other analyses of tacit beliefs which could make better sense of basic normative beliefs, but since I am not aware of any such analyses I shall simply leave it at that.

### **3.5 Internal reasons and propositional attitudes holism**

Since basic normative beliefs cannot plausibly be thought of as tacit beliefs, if one nonetheless wants to retain the idea of basic normative beliefs, perhaps we should take more seriously the idea that there can be a deflationary account of basic normative beliefs. To flesh out this idea, we should begin by noticing that in the passage concerning the possibility of beliefs being motives, Williams

seems to make at least an implicit appeal to a holistic picture of propositional attitudes. Very roughly, holists about propositional attitudes hold that propositional attitudes such as beliefs and desires do not, and cannot, exist 'in a vacuum'. As Davidson puts it:

Propositional attitudes, in terms of which mental events are characterised, cannot exist in isolation. Individual beliefs, intentions, doubts and desires owe their identities in part to their position in a large network of further attitudes: the character of a given belief depends on endless other beliefs; beliefs have the role they do because of their relations to desires and intentions and perceptions. These relations among the attitudes are essentially logical: the content of an attitude cannot be divorced from what it entails and what is entailed by it.<sup>16</sup>

I think we can separate two different claims in this passage: one metaphysical, the other normative (in a quasi-semantic/epistemological sense). The normative constraint on the correct *attribution* of attitudes should, I think, be taken as an epistemological (or quasi-semantic) constraint: we cannot *legitimately attribute* a certain mental state to a person without attributing certain other mental states to that person:

There is no assigning beliefs to a person one by one on the basis of his verbal behaviour, his choices, or other local signs no matter how plain and evident, for we make sense of particular beliefs only as they cohere with other beliefs, with preferences, with intentions, hopes, fears, expectations, and the rest.<sup>17</sup>

Furthermore, according to Davidson, the legitimate attribution of one or more propositional attitudes to another person also requires us to not only attribute other beliefs and desires to that person, but also to do this in accordance with what Davidson has called the 'principle of charity'. Roughly, this principle bids us to minimise the ascription of false beliefs to a person. Although this principle is of great importance to the project of 'radical translation' it is of no direct concern to us here.

Putting these issues aside, we must now consider what some of the other implications of holism are. Davidson says that beliefs and desires *cannot exist* independently of other, 'related' beliefs and desires. This would, if true, rule out the possibility of a person's having, for instance, only one belief. This certainly seems true. However, Davidson thinks that the interdependency of propositional attitudes is a *logical* one ('the content of an attitude cannot be divorced from what it entails and what is entailed by it'). I take it

that he does not mean by this that a person who, say, believes  $p$  must believe all that  $p$  entails. This would be to assume too much. For instance, if we take every true mathematical proposition to entail all other true mathematical propositions, then if I believe a true mathematical proposition, it would follow that I believe all true mathematical propositions. But clearly this is not the case (except perhaps if I believe these propositions tacitly). We should therefore think of Davidson's idea as being less ambitious than that. The idea might simply be that I cannot have the belief that, say,  $3 + 3 = 6$  without having *some* other belief about mathematics and/or arithmetic. Likewise, John could not believe that a burglar just entered his house through a window without believing such things as 'burglars exist' or 'there are windows in my house'. Again, John need not have *these* particular beliefs; the point is simply that he cannot have the belief that a burglar has just entered his house through a window without having *some* other, 'related', beliefs. As I understand holism it is indeterminate what other beliefs John *has* to have in order to have the one just ascribed to him. But this indeterminacy is no embarrassment to holism. It simply says that a necessary condition for the obtaining of a belief (or any other propositional attitude) is that this belief belong to a network of other, interrelated beliefs.

However, and more importantly for our purposes, the interrelatedness of propositional attitudes can, and does, 'reach

across', so to speak, different attitude types - e.g. from beliefs to desires. This picture is further supported by common-sense psychology. As a theory, common-sense psychology consists in a body of generalisations relating psychological states to each other, to inputs from the environment, and to actions. Here is a sample from Churchland:

(x) (p) (if x fears that p, then x desires that not-p)

(x) (p) (if x hopes that p and x discovers that p, then x is pleased that p)

(x) (p) (q) (if x believes that p and x believes that if p then q, then barring confusion, distraction etc. x believes that q)

(x) (p) (q) (if x desires that p and x believes that if q then p, and x is able to bring it about that q, then barring conflicting desires or preferred strategies, x brings it about that q.<sup>18</sup>

All of these generalisations should be understood as containing *ceteris paribus* clauses, but the similarity with Williams' picture is obvious enough:

A man who does believe that considerations of family honour constitute reasons for action is [barring akrasia] a



man with a certain disposition to action, and also dispositions of approval, sentiment, emotional reaction, and so forth.<sup>19</sup>

What emerges from these considerations is, I think, the idea that a person who believes that he has a reason to do that which family tradition bids him to do, does so because he in some way or other desires to do that which family tradition bids him to do. In so far as Owen's belief about what family tradition bids him to do (i.e. to join the army) is correct, his belief that this gives him a reason to join the army is made true by the fact that he has a basic desire to do that which family tradition bids him to do.

It should of course be admitted that no such claim (concerning desires) is being made in the passage quoted above. At best, what we are entitled to conclude from the preceding considerations (in conjunction with the quote from Williams) is that since it cannot be Owen's belief (or its content) that provides him with a reason to join the army, it must be his dispositions that somehow give him a reason to join the army. However, it is hard to see how a person's dispositions can be reason giving. The most natural way of interpreting Williams here is to read him as saying that a man who has the kinds of beliefs under consideration is a man with certain desires (which, as I suggested in the previous chapter, should be taken to be dispositional states).

On this account then, if Owen believes that he has a (basic) reason to do whatever family tradition bids him to do, it will be true of Owen that he does have a basic reason to do whatever family tradition bids him to do. This is consistent with (16). And this may be enough to legitimately call Owen's belief basic. However, on the view under consideration, Owen's belief is not made true by his believing it; rather, it is made true by his having a basic desire to do that which family tradition bids him to do. In other words, if Owen's belief that he has a reason to  $\phi$  is basic, then Owen has some basic desire to  $\phi$  which guarantees the truth of his belief (according to internalism). A consequence of these considerations is that we could have said right from the start that Owen's belief that  $p$  is made true not by his believing that  $q$  but by his basic desire to do whatever family tradition bids him to do; a desire which finds an expression in his belief that he has reason to do that which family tradition bids him to do. So we can still legitimately say of Owen that his belief that  $q$  is a basic normative belief while holding that to have a basic normative belief is nothing over and above having a certain basic desire. Furthermore, the idea of tacit beliefs does not seem entirely misplaced here. Perhaps we should say of Owen that his basic normative belief that  $q$  is a tacit belief which we can legitimately attribute to him in virtue of his have a basic desire to do whatever family tradition bids him to do.

In the end, then, the more plausible alternative for the internalist is to abandon (11) altogether in favour of (12). On this line of thought, the element of Owen's motivational set in virtue of which his belief that  $p$  is true is his desire to do what family tradition bids him to do. In other words, the reason why (4) is true is that Owen desires to do that which family tradition bids him to do. But if this is correct it is hard to see what significance Owen's belief that  $p$  has in the 'reason-generating' process. If it is Owen's *desire* to do whatever his family tradition bids him to do that is the ultimate source of his having a reason to join the army, then he would presumably have had this reason even if he had not had the belief that  $p$ . But then, of course, it is false, or at least grossly misleading, to say that Owen has a reason to join the army *because* he believes that the fact that family tradition bids him to do so provides him with a reason to do so. What is *really* doing all the work here is his desire to honour family tradition. So the simplest and, in the end, correct picture is that beliefs do not generate reasons, desires do. This is of course assuming that Owen's desire to  $\psi$  is not based on a belief. If it is, then we can run the same argument all over again. Therefore, Williams is forced to say that what we have reason to do is ultimately grounded in some basic desire (i.e. one that is not based on belief). These basic desires, in turn, are ones we have no reason to have and which we also have no reason not to have. If Owen's desire (or motivation) to  $\psi$  came

about because he believes himself to have a reason to  $\psi$ , it follows that there must be some other, even more basic desire in virtue of which his belief that he has reason to  $\psi$  would be true. Therefore, since this cannot be the case on pain of regress, Owen's desire to  $\psi$  (or something like this desire) must be basic.

According to this picture, the answer to Williams' question 'Does believing that a particular consideration is a reason to act in a particular way provide, or indeed constitute, a motivation to act?' might be 'yes', but as it turns out, Owen's belief that  $p$  is merely a *consequence* of having the desire to  $\psi$ , which in turn is the *real* source of Owen's reason to join the army (if indeed he has such a reason). So although, as Williams puts it, 'this agent, with this belief, appears to be one about whom, now, an *internal* reason statement could truly be made: he is one with an appropriate motivation in his  $S'$ , it is not *because* he has that belief that we can make a true reasons statement about him. The inclusion of 'now' in the passage above seems to suggest that before Owen had the belief that  $p$ , it was simply false that he had a reason to join the army. We can now see that this is not the case. It became true of Owen (on this picture) that he has a reason to join the army when he formed the desire to do whatever family tradition bids him to do.

### 3.6 Desire-based reasons

Given the preceding considerations, I think we should conclude that internalists are committed to the claim that only desires (or desire-like states perhaps) are 'ultimate' motives. So if we have reason to  $\phi$  only in so far as we have a motive which either consists in, or which would be served by our  $\phi$ -ing, then since, according to internalism, motives simply are desires, we have reason to  $\phi$  only in so far as we have a desire which either consists in, or which would be served by our  $\phi$ -ing. Setting aside the issue of the supposed explanatory dimension of normative reasons, and keeping in mind the proviso that desires based on false beliefs are not reason giving, we can say that on the internalist conception of reasons, reasons are desire-based.

There are, broadly speaking, two kinds of desires, instrumental desires, and intrinsic (or 'basic') desires. In a sense, instrumental desires are rationalised by other desires in a way that intrinsic desires are not. As Parfit puts it:

We often have long chains of instrumental desires, but such chains all end with some intrinsic desire. Thus, we may want medical treatment, not for its own sake, but only to restore our health, and we may want that, not for its own sake, but only so that we can finish some great work of art, and we may want that, not for its own sake, but

only to achieve posthumous fame. This desire may in turn be instrumental, since we may want such fame only to confound our critics, or to increase the income of our heirs. But, if we want posthumous fame for its own sake, this intrinsic desire would end this particular chain.<sup>20</sup>

These intrinsic desires, then, are, on the internalist conception of reasons, what ultimately grounds our reasons for action.

Now an important feature of internalism (construed in the way we are currently considering) is that our reasons for action are ultimately grounded in our intrinsic desires which, in turn, we have no reason to have and no reason not to have; because if we did have a reason either for or against having a particular basic desire - the desire to listen to a particular piece of music, say - then since all reasons are supposedly grounded in our desires, the desire to listen to this particular piece of music cannot be a basic desire. In Hume's terminology, a basic desire is an 'original existence' - either we have it or we don't, and there can be no reason either for or against having that particular desire. As Hume puts it: "Tis impossible...that...passion can be oppos'd by, or be contradictory to truth and reason..."<sup>21</sup> I think this view is mistaken. To see why this view is mistaken, let us make a slight detour and take a look at a very closely related issue: the role of desires in motivation.

According to the standard interpretation of Hume's theory of motivation, beliefs are, by themselves, incapable of motivating us to act. Without going too deeply into Humean exegesis and to avoid the possibility of misrepresenting Hume, we can simply call the theory that beliefs are by themselves incapable of motivating the 'Humean theory of motivation'. According to this theory, all intentional actions are the result of the interaction of two distinct mental states: a cognitive state and a non-cognitive state. Many contemporary philosophers have found it difficult to draw a definitive line between these kinds of states. However, they do agree that beliefs and desires are paradigm cases of cognitive states and non-cognitive states, respectively. For the Humean, then, no beliefs are by themselves capable of motivating; only when keyed to a suitable desire will beliefs motivate.

Many philosophers today reject the Humean theory of motivation, and, of those who do, many often cite Thomas Nagel's distinction between *motivated* and *unmotivated* desires as their reason for doing so. Without going into too much detail, Nagel's account of motivation is compatible with the Humean account in so far as both state that motivation involves the existence of a 'suitable set' of beliefs and desires. But, says Nagel, if we allow that there are such things as motivated desires, the significance of desires in motivation is significantly diminished<sup>22</sup>:



The claim that a desire underlies every act is true only if desires are taken to include motivated as well as unmotivated desires, and it is true only in the sense that *whatever* may be the motivation of someone's intentional pursuit of a goal, it becomes in virtue of his pursuit *ipso facto* appropriate to ascribe to him a desire for that goal. But if the desire is a motivated one, the explanation of it will be the same as the explanation of his pursuit, and it is by no means obvious that a desire must enter into this further explanation.... That I have the appropriate desire simply *follows* from the fact that these considerations motivate me....<sup>23</sup>

So Nagel agrees that being motivated involves both beliefs and desires, but desires play no *crucial* causal role in the story; the resulting desire is a *consequence* of having the appropriate belief. Jonathan Dancy has recently pointed out that there are at least two possible interpretations of this passage of Nagel's:

The first one allows there to be a desire present, as an independent existence (in Hume's sense), but insists that it is not playing a Humean role. It is motivated, not motivating; we know that it is there because though we know that the beliefs in the case are sufficient explanation

of the action, we allow Hume the point that every action is caused by a complex which includes a desire. The desire which must be present, then, must be one which is explained by the beliefs and which does not explain the beliefs' ability to motivate. The second theory maintains that though we *ascribe* a desire to a person motivated entirely by his beliefs, for instance a care for his own physical well-being, all that is meant by this is just that his beliefs were sufficient reason for him to act. Here we do not admit the need for a desire as an independent existence; we *call* the belief's motivating the agent 'his doing it because he wanted'.<sup>24</sup>

For our purposes we don't need to decide which one of these interpretations is the correct one (perhaps they are both correct). But why is the Humean theory of motivation relevant to the question of whether or not desires are reason-giving? The answer is that Nagel has provided us with a very good reason for believing that at least some non-instrumental desires are not 'original existences' in Hume's sense. Consider the following passage from Hume:

Ask a man *why he uses exercises*; he will answer *because he desires to keep his health*. If you then enquire *why he*

*desires health*, he will readily reply *because sickness is painful*. If you push your enquiries further and desire a reason *why he hates pain*, it is impossible he can ever give any. *This is an ultimate end*, and is never referred to any other object...beyond this it is an absurdity to ask for a reason. It is impossible there can be a progress *in infinitum*; and that one thing can always be a reason why another is desired. Something must be desirable on its own account....<sup>25</sup>

I think this passage is significant since it reveals, at least in part, the real structure of our desires (even if Hume himself would have denied it). As Parfit puts it:

It is worth noting how, when Hume described such a chain of instrumental desires, he forgot his own theory. ... For 'desirable' Hume should have written 'desired'. Something is desirable if it has features that give us reasons to want this thing. Hume denied that there could be such reasons.<sup>26</sup>

Parfit's observation is, I believe, correct. However, this does not mean that anyone who wants to oppose Hume is committed to the claim that all desire chains must terminate in some object which is

desirable in itself. If this were so, then opponents of Hume would have to say that our desires are infallible in the sense that they all track what is desirable, and clearly this is not so. What opponents of the Humean picture should say instead is that our desire chains must terminate in a *belief* or *judgement* that something is desirable in its own right; a belief which, of course, can be either true or false. So although Hume's own account of desires, does not allow him to say that some objects are desirable in their own right (or that we judge them to be so), Hume may (perhaps unwittingly) be expressing something almost all of us take to be true: unless we thought that some action or object has some desirable features (i.e. that it has properties such that they provide us with reasons to desire that object), it would not make much sense to desire the thing in question.

To see that this is so, imagine someone, S, who professes to desire something for its own sake but who also vehemently (and honestly) denies that he takes this thing to be *desirable*; i.e. he denies that there is anything good about this thing. Consider the following example from Scanlon (which he in turn has borrowed from Quinn):

[Consider] a man who feels an urge to turn on every radio he sees. It is not that he sees anything *good* about radios being turned on; he does not want to hear music or news

or even just to avoid silence; he simply is moved to turn any radio he sees to be off.<sup>27</sup>

How could we make sense of such a person? There is certainly something odd about saying that this person *desires* to turn on radios. Rather, this person appears to be someone who is suffering from some kind of obsessive-compulsive disorder. It seems that part of desiring something includes an evaluative aspect (i.e. seeing something good in, or about, the thing which is desired), and if this component is absent from a purported desire, it is more appropriate, I think, to describe this 'volition' as merely an urge.

I think we should say that all those desires which Nagel called 'unmotivated', are better understood as urges. If an agent is moved to do something without seeing anything good or worthwhile about doing this thing, then although this would be an example of an unmotivated desire in Nagel's sense, it is somewhat misleading to call it a desire. I want to suggest that all unmotivated desires are, simply, urges. But if an agent is or would be moved to  $\phi$  because of an urge he has (or would have), it is hard to see why we should say that he has a reason to  $\phi$  - even if that urge is not the product of a false belief. After all, the agent himself admits that he sees nothing good or worthwhile about  $\phi$ -ing. It seems reasonable to suppose that such an agent would *agree* that he has no reason to

do that which he has the urge to do - and I think that, all things being equal, he would be right about this.

In the end, then, I think we arrive at the following picture: when we 'desire' an object, an action, or a state of affairs, we do so either because we see something good or worthwhile about that object, action, or state of affairs, or we do so because we simply have an urge to bring about the object of our urge. If someone is moved purely by an urge, I think we should say that, all things being equal, this person has no reason to do that which his urge bids him to do. On the other hand, if an agent's desire is motivated by some perceived good in or about an object or an action, then whether or not this agent has a reason to do the thing in question depends on whether his belief (perception) about the goodness of this object, action, or state of affairs is true. However, as Raz points out, the desire in this case is an endorsement of a reason (or a perceived reason) that is independent of the desire, and that endorsement does not itself affect the stock of reasons that existed before the agent came to endorse or appreciate them.<sup>28</sup>

### **3.7 Desires and instrumental reasons**

I think the preceding considerations provide us with the sufficient resources to dispel the Humean myth that 'Reason is, and ought only be the slave of the passions'.<sup>29</sup> According to the Humean

picture, since our 'ultimate' desires are not (and, indeed, cannot) be based on Reason (or reasons), Reason's only job is to point out the various ways in which our ultimate (and derivative) desires can be satisfied. But, as Williams points out: 'the mere discovery that some course of action is the causal means to an end is not in itself a piece of practical reasoning.'<sup>30</sup> So, on the one hand, Hume's claim can be interpreted as implying that there can be no such thing as practical reason *at all*. On the other hand, Hume's claim can be, and has been, interpreted as saying that, for any basic desire to  $\phi$ , although we can have no reason to have that desire and no reason not to have that desire, if we do have a basic desire to  $\phi$ , and if  $\psi$ -ing is suitably related to  $\phi$ -ing (e.g. by being a necessary means to the satisfaction of our desire to  $\phi$ ), then we do have a reason to  $\psi$ . I think there are good grounds for believing that Hume himself held the former view; but since it is common place for Humeans to assert the latter, I want to say something about this view.

According to these Humeans, then, although we can have no reason to have the basic desires we in fact have, we do have reason to act in certain ways in virtue of having those basic desires. The view can be represented schematically as follows:

If A has a basic desire to  $\phi$  and if A's  $\psi$ -ing is suitably related to A's  $\phi$ -ing, then A has a reason to  $\psi$ .<sup>31</sup>



I think we should reject the Humean schema and the philosophical picture that underlies it. I have two closely related arguments against this view. First, consider the analogous case of beliefs. Suppose A believes that  $p$ , and A believes that if  $p$  then  $q$ , but A has no reason for believing  $p$ . Here we might be tempted to think that A has, in virtue of the beliefs he has, at least some reason to believe  $q$  since  $q$  obviously follows from the contents of the beliefs A has. But this is a mistake. To see this, imagine that A's belief that  $p$  is a contradiction and that A believes (correctly according to at least some logicians) that everything follows from a contradiction. On the current view this would imply that for every proposition there is (true and false) A has a reason to believe it. Clearly this is not so.

It might be objected that the analogy between the practical case and the theoretical one is illegitimate. I shall not try to defend the analogy. For those who are not persuaded I will offer the second argument (which is due to Jonathan Dancy). It is a very simple argument, consisting of only one premise and a conclusion:

*Premise:* A desire to  $\phi$  cannot itself give us any reason to  $\phi$ .

For if  $\phi$ -ing is silly or even just not very sensible, wanting to  $\phi$  does not make it less silly or a bit more sensible.

*Conclusion:* If a desire to  $\phi$  gives us no reason to  $\phi$ , it can give us no reason to do other actions either; in particular, it can give us no reason to do those actions that subserve  $\phi$ -ing (either as a means to  $\phi$ -ing as end, or in some other way).<sup>32</sup>

Both these arguments are, in effect, different versions of Michael Bratman's 'bootstrapping' objection<sup>33</sup> which in turn can be seen as a version of Hume's own *No 'ought' from an 'is'* principle (where, of course, we need to replace 'ought' with 'reason(s)'). The main idea behind both these arguments is that unless we already have a reason to  $\phi$  or a reason to believe that  $p$  and to believe that if  $p$  then  $q$ , we cannot have a reason to  $\psi$  or a reason to believe that  $q$  simply in virtue of having certain other desires or beliefs. In other words, we cannot bootstrap a reason into existence from nowhere.

For those who are still not persuaded - i.e. for those who think that if we have a basic desire to  $\phi$ , or if we believe that  $p$  and we believe that if  $p$  then  $q$ , then in some sense, we must have some reason to  $\psi$  and some reason to believe that  $q$ , respectively - let me say something about a very persuasive proposal put forward by John Broome.<sup>34</sup> Broome agrees with Bratman that it is a mistake to think that we can bootstrap reasons into existence from our beliefs and desires independently of whether these are already 'normatively sanctioned' in some way. In the following short

exposition of Broome's argument I shall follow him in talking about intentions rather than desires or beliefs although the same argument can be applied to those attitudes as well. Consider the following principle:

- (I)            If A intends to  $\phi$  and A believes that  $\psi$ -ing is a necessary means to  $\phi$ , then A has a reason to intend to  $\psi$ .

This principle has a 'detachable' consequent; i.e. the proposition 'A has a reason to intend to  $\psi$ ' can be validly detached or derived from (I) if A satisfies the conditions in the antecedent. However, says Broome, this is an illegitimate move; intending to  $\phi$  and believing that  $\psi$ -ing is a necessary means to  $\phi$  cannot give you a reason to intend to  $\psi$  even if the belief that  $\psi$ -ing is a necessary means to  $\phi$ -ing is true. A defender of (I) might say that A has a reason (or even more strongly, that he ought) on pain of inconsistency to intend to  $\psi$ . But, says Broome, this is not so. The consistency requirement is correctly captured by the following principle which he calls a 'normative requirement':

- (NR)            A has a reason to (intend to  $\psi$  if A intends to  $\phi$  and A believes that  $\psi$ -ing is a necessary means to  $\phi$ ).<sup>35</sup>

Broome's point, then, is that although we might think that a person who has certain beliefs, desires, or intentions, has at least *some* reason to believe, desire, or intend that which those attitudes seem to call for on pain of inconsistency, this is not so. Rather, we are normatively required to believe (at least the immediate) consequences of our beliefs; to desire to take the necessary means to that which we desire; and to intend the necessary means to that which we intend. This does not, however, mean that we have a reason to believe, desire, or intend these things. As he puts it:

Relations among your beliefs and intentions are regulated by oughts that govern these relations - 'wide-scope oughts' I shall call them. They imply no narrow-scope normative conditions on individual beliefs or intentions. ... Intending an end clearly stands in some sort of normative relation to intending a means. So if the only normative relation you think of is the relation of being a reason to, you are likely to think that intending an end is a reason to intend a means. Then your view implies that, if you intend an end, you have a reason to intend a means. That is to say, if you intend an end, then the narrow scope normative condition of 'having a reason to' applies to your intending a means.

But this is to misunderstand that structure of normativity. Suppose you intend to visit Rum, and the only way you can get there is to take the boat. A reason to intend to take the boat is that it will carry you to the wild and beautiful island of Rum. This reason exists independently of your intention to visit Rum. But if this intention was also a further reason to intend to take the boat, it would be a reason you create yourself by forming the intention to visit Rum. It is puzzling how you could create such a reason in that way....<sup>36</sup>

It is obvious from this passage that Broome is expressing an externalist view of reasons. For present purposes, however, this makes no difference. I introduced Broome's views on normative requirements for the purpose of convincing those who believe that consistency requirements give us reasons to, e.g., believe the immediate consequences of what we believe. By accepting Broome's normative requirements we can accommodate our intuition that we ought to be consistent in such a way that we can retain that intuition without thereby committing ourselves to the idea that our beliefs, desires, or intentions are, in themselves, reason giving.

### 3.8 Beliefs as sources of motivation

I have argued in this chapter that Williams cannot plausibly hold that beliefs can be motives while at the same time retain his internalism about reasons. I have tried to show that Williams is, in the end, committed to a Humean, or quasi-Humean view of reasons: the reasons we have for acting are given by our desires; desires which, in turn, we have no reason to have and no reason not to have. I think we should reject the Humean view and, consequently, internalism about reasons. To set the stage for the discussion about externalism which will follow in the next chapter, I want to say something briefly about how and why we should accept the thesis that beliefs are motives and that beliefs can motivate by themselves.

Earlier I said that we should think of a motive as that which can be adduced in our everyday explanation of intentional actions. So if beliefs can be motives, it must be possible that we could explain a person's intentional action by referring to that person's beliefs alone. For *everyday* explanations of action, I think citing a person's beliefs is indeed sufficient, but a more robust philosophical explanation of action requires more than this.

Suppose we were to ask why Bob went to Glasgow yesterday. The anti-Humean might respond by saying that Bob went to Glasgow because he believed that by doing so he could watch Hibs play Rangers at Ibrox. The Humean would undoubtedly respond by

saying that the fact that Bob believed that he could watch Hibs play Rangers in Glasgow is at best an incomplete explanation of his action. A correct and complete explanation would have to include not only the fact that Bob believed that Hibs were playing Rangers at Ibrox, but it would also have to include some desire on Bob's part to attend the match. The Humean might argue along the following lines: Suppose Bob had no desire whatsoever to watch Hibs play Rangers, but he nonetheless believed that Hibs were playing Rangers in Glasgow. How then could the fact that Bob believed that Hibs were playing Rangers explain his going to Glasgow yesterday? Surely we need to include some desire of Bob's in a correct explanation of his going to Glasgow?

There are, as I see it, a number of answers available to the anti-Humean here. First, the anti-Humean should concede to the Humean that the fact that Bob believed that Hibs were playing Rangers at Ibrox is not, by itself, sufficient to explain Bob's going to Glasgow; Bob must also have had some kind of 'evaluative' belief to the effect that it would be entertaining or worthwhile for him to watch the match. However, since to believe that something is worthwhile is to believe that there are reasons for doing the thing in question (more on which in the next chapter), we can say that Bob must have had some belief to the effect that he had good reason to attend the match. Bob's belief that Hibs were playing Rangers at Ibrox (as opposed to at Easter Road in Edinburgh) should thus be



seen merely as a belief about the necessary means to 'promote' or 'realise' that which he believed to be worthwhile (or that which he took himself to have good reason to do) - namely, to attend the match.

The Humean might at this point simply repeat his challenge by insisting that this only shows that Bob must have a desire to do that which he finds worthwhile (or that which he takes himself to have good reason to do). In effect, this Humean challenge is simply the practical analogue of the tortoise's challenge to Achilles in Lewis Carroll's famous story.<sup>37</sup> In Carroll's original story the tortoise challenges Achilles to show why someone who refuses to conclude 'q' from 'if p then q' and 'p' is making a mistake. The tortoise suggests to Achilles that the conclusion follows only if the inference from (If [if p then q and p] then q) is valid. Achilles agrees that this is the case, but then to his dismay he realises that he has just embarked on an infinite regress. The moral of the story is that Achilles cannot show that someone who refuses to make this inference is making a mistake without invoking the very thing the tortoise wants to challenge - namely the validity of Modus Ponens. This story has recently been modified by Simon Blackburn who presents it as a challenge to anti-Humeans in the theory of motivation. The moral of Blackburn's story is that

there is always something else, something which is not under the control of fact or reason, which has to be given as a brute extra, if deliberation is ever to end by determining the will.<sup>38</sup>

The 'brute extra' Blackburn is referring to here is of course an unmotivated desire. The challenge to the anti-Humean, then, consists in his showing that this sort of regress can be stopped without invoking a desire as the terminating 'brute extra'. I think the anti-Humean can meet this challenge. To begin with, the anti-Humean must obviously reject the idea that Bob's belief that it would be worthwhile to attend the match really is a desire in disguise. This does not mean that the anti-Humean must deny that Bob's belief that attending the Rangers v. Hibs match is a worthwhile activity motivated him to travel to Glasgow. If the Humean insists that a desire must be included in the explanation of Bob's going to Glasgow, the anti-Humean can quite happily admit that it is perfectly legitimate to ascribe to Bob such a desire simply in virtue of Bob's being motivated to do so. Crucially, however, this desire is not independent of Bob's belief that it would be worthwhile to travel to Glasgow. On the contrary, Bob's desire to travel to Glasgow would be unintelligible without his belief that it is a worthwhile thing to do.

Of course, for it to be true of Bob that his belief motivated him, it must be true of Bob that he has a disposition such that (at least some of) his evaluative or normative beliefs motivate him. As Parfit puts it (and to repeat a point made in chapter one):

When we come to have some belief - such as the belief that some aim is worth achieving - that might cause us to have some wholly new desire. Such a belief could not all by itself cause us to have this desire, since we would have to be *such that*, if we came to have this belief, that would cause us to have this desire. But this disposition may not itself be a desire. On a variant of this anti-Humean view, whenever a belief moves us to act, we can be truly said to have wanted to act as we did; but this desire may not be a distinct mental state, since it may consist in our being moved by this belief. In either of these ways, reason might have the power that Hume denied. By giving us such beliefs, reason might motivate us without the help of any independent desire.<sup>39</sup>

This disposition - being such that if we have certain normative and/or evaluative beliefs we are then motivated to act on these beliefs - seems to be exactly the kind of 'brute extra' the anti-Humean needs to answer Blackburn's challenge. This brings us

back to the discussion of rationality in chapter one. As we can now see, the disposition the Humean has to invoke is nothing over and above the very intelligible and non-mysterious disposition of acting according to one's normative beliefs, i.e. being rational in at least the formal sense (as described in chapter one).

I think these considerations strongly suggest that we can and should reject the Humean picture of motivation. However, I want to consider one more objection to the broadly cognitive view of motivation I have suggested we should accept. This objection is also related to issues we have already discussed. It stems from the observation that an agent might have various evaluative or normative beliefs without being motivated to act on those beliefs - if he is acciduous or weak willed for example. Suppose someone who believes that it would be good (in some sense or other) to attend this afternoon's lecture fails to do because he is weak willed or acciduous. Doesn't this simply show that this person lacked the requisite desire to go to the lecture? And if the absence of an appropriate desire in this case is sufficient for explaining why this person didn't attend the lecture, doesn't this show that was right after all: motivation (and action) can never take place without the presence of suitable desires.

The correct answer to this objection consists in holding that the disposition to act in accordance with one's normative or evaluative beliefs is such that it doesn't issue in actions under

certain conditions. Since we can fail to act in accordance with our normative or evaluative beliefs, the disposition to act in accordance with our normative or evaluative beliefs must be understood as containing a 'barring weakness of will and accidie etc.' clause. A person who is rational in the formal sense is a person who acts in accordance with his or her normative beliefs *under normal circumstances*. However, when circumstances are not normal there will some explanation of why this counts as a non-normal circumstance. As Dancy points out, there are

no restrictions of the sorts of explanation that we are prepared to countenance. Sometimes the reason will be carelessness or inattention; sometimes it will be despair; sometimes it will be an excess of alcohol; sometimes it will be a neuro-physiological disorder; and sometimes it will be clinical depression.<sup>40</sup>

I think we should say, then, for a person who is rational in the formal sense, his or her normative beliefs are capable of motivating in their own right. But when they fail to do so there will be an explanation of this fact; an explanation which will proceed via an analysis of what it means to be rational. Normative or evaluative beliefs are, as Dancy puts it, 'contingently, but intrinsically motivating states'.

### 3.9 Summary and conclusion

In this chapter I have argued that internalist cannot plausibly allow that beliefs can be motivating states. In general, propositions of the form 'The fact that  $p$  gives  $A$  a reason to  $\phi$ ' (which is the kind of proposition Williams considers) must be construed as external reasons statements. And as such, they cannot ground internal reasons statements - even if they motivate - on pain of contradiction. Even if a proposition of the form just mentioned is true on an internalist reading - i.e. even if ' $p$ ' refers to what  $A$  is or would be motivated to do - and even if such a proposition motivates an agent, it is hardly going to be *made true* by its motivating power (if it has any). Rather, such propositions are more likely to be mere restatements of the internalist thesis itself. Internalism is plausible only on a Humean theory of motivation. I have provided some arguments against this theory. We should reject internalism and Humeanism. In the next chapter I shall consider various externalist answers to the questions 'why should I be moral?'.

---

<sup>1</sup>Williams (2001).

<sup>2</sup>Skorupski (forthcoming).

<sup>3</sup>Williams (1981), p. 107.

<sup>4</sup>Throughout this reconstruction I am going to assume (with Williams) that the agent - in this case Owen - is not akratic.

<sup>5</sup>If this is correct, then we should reject what I called reason judgement/motivation internalism in the previous chapter.

<sup>6</sup>It might be objected that this is a false dichotomy. It is possible to hold that the feature of O's S in virtue of which p is true is a 'besire' - a mental state which has both the 'direction of fit' of a belief and the direction of fit of a desire (see Altham (1986), McDowell (1978), McNaughton (1988), and Smith (1994)). I must however leave this point to the side.

<sup>7</sup>Williams (1981), p. 102.

<sup>8</sup>See Lycan. I should point out that my discussion of tacit beliefs is heavily informed by Lycan's discussion.

<sup>9</sup>See Field.

<sup>10</sup>I have borrowed these examples from Lycan, p. 61.

<sup>11</sup>Lycan, p. 65.

<sup>12</sup>*Ibid.*, p. 66. There seem to be some obvious problems with this approach. I take it that Dennett means by 'consequences of those beliefs' those propositions which are implied by the content of that belief. If that is so, suppose that A has somehow stored two contradictory beliefs. If A's extrapolator-deducer is in good working order, then since everything is implied by a contradiction, it should generate...what? Every proposition there is? But clearly nobody in their right mind would want to say that such an agent tacitly believes everything. An obvious way of getting round this problem would be to stipulate that one cannot have contradictory *core* beliefs. This move obviously needs to be supplemented by some story about what can count as a core belief and what cannot. Suppose that one of A's core beliefs is that all extended objects have mass. It is an obvious 'consequence' of this belief that all extended objects have mass or Tottenham play attractive football. But what if A would not judge this to be so? Does that mean that he does not have the belief that all extended objects have mass; or is it perhaps the case that he does have this belief, only it is not a core one; or, again, perhaps he doesn't understand the rule of 'or-introduction'?

<sup>13</sup>We can see now that even if we allow that Owen's basic belief is true, his derivative belief, as it stands is at best incomplete and at worst false. It is not the case that [the fact that family tradition bids him to join the army] gives him a reason to join the army. Rather it is the conjunctive fact that [he has a reason to do



whatever family tradition bids him to do and family tradition bids him to join the army] which gives him a reason to join the army.

<sup>14</sup>See Field.

<sup>15</sup>Lycan, p. 70.

<sup>16</sup>Davidson (1994), p. 232.

<sup>17</sup>Davidson (2001), p. 221.

<sup>18</sup>Segal, p. 146.

<sup>19</sup>Williams (1981), p. 107.

<sup>20</sup>Parfit (forthcoming), see Ch. 6.

<sup>21</sup>Hume (1978), p. 415.

<sup>22</sup>It should perhaps be pointed out that although Nagel's discussion of the possibility of a purely cognitive theory of motivation is mainly concerned with prudential considerations, other authors, especially McDowell, have pressed the need for extending this theory of motivation to cover moral considerations as well.

<sup>23</sup>Nagel, p. 29.

<sup>24</sup>Dancy (1993), pp. 8-9.

<sup>25</sup>Hume (1975), p. 293.

<sup>26</sup>Parfit (forthcoming), Ch 3.

<sup>27</sup>Scanlon (1998), p. 38.

<sup>28</sup>Raz (1986), p. 141.

<sup>29</sup>Hume (1978), p. 415.

<sup>30</sup>Williams (1981), p. 104.

<sup>31</sup>What should count as 'suitable' here is a matter of controversy. Most, if not all, Humeans agree that there is a suitable relation between  $\phi$ -ing and  $\psi$ -ing if the latter is a necessary means to the former. But this does not mean that 'being a necessary means to' is the only suitable relation. As Williams says 'A clear example of practical reasoning is that leading to the conclusion that one has reason to  $\phi$  because  $\phi$ -ing would be the most convenient, economical, pleasant etc. way of satisfying some element in S....' (Williams (1981), p. 104.)

<sup>32</sup>Dancy (2000), p. 32.

<sup>33</sup>Bratman, pp. 23-7. See also Broome (1997) and (2001).

<sup>34</sup>Broome (1997).

<sup>35</sup>More precisely, Broome calls this a 'normative recommendation'. This is a different normative relation from normative requirements which have the stronger 'ought' operator governing the conditional. The difference need not concern us here.

<sup>36</sup>Broome (forthcoming).

<sup>37</sup>Carroll, pp. 278-80.

<sup>38</sup>Blackburn, pp. 695-711.

<sup>39</sup>Parfit (1997), p. 105. On an alternative view, one could hold that there's no need to posit such a disposition. On this view (call it dispositionalism), all mental states, including beliefs, are dispositional. According to this view, A's believing that p consists in

---

(among other things) A's being disposed (motivated) to infer  $q$  if A believes if  $p$  then  $q$ ; A's being motivated to give up his belief that  $p$  if presented with good evidence for not- $p$ ; A's being motivated to assert  $p$  if sincere etc. Although I think there's quite a lot to be said for this view, I shall not defend it here.

<sup>40</sup>Dancy (1993), p. 25.

## **CHAPTER FOUR: EXTERNALISM AND WHY WE SHOULD BE MORAL**

### **4.1 Introduction**

In the previous two chapters I provided some arguments against internalism about reasons for action. I do not want to suggest that these arguments are conclusive but I nevertheless think they put enough pressure on internalism to make it unattractive as an analysis of reasons. It is important to stress that I do not think we should reject internalism (merely) because it gives a particular answer to the question 'why should I be moral?' (an answer which, incidentally, I have not confronted directly, though it's general shape should be fairly obvious) - rather, we should reject internalism because it provides us with an implausible, and I would add, mistaken, account of reasons and normativity. Since I think we should reject internalism, I think we should be externalists about reasons. But what exactly does externalism about reasons amount to? And more importantly, how does externalism about reasons relate to the question we are interested in - 'Why should I be moral?' In this chapter I will try to sketch at least a preliminary answer these two questions.

## 4.2 Externalism, reasons, and value

So far I have said only that externalism denies the central claim of internalism – the claim that a necessary condition for the obtaining of a reason for  $A$  to  $\phi$  is that  $A$  either has a motive which will be served by his  $\phi$ -ing or that  $A$  would in suitable circumstances be motivated to  $\phi$ . There is however a slight complication here.

Strictly speaking, externalists need not deny that if  $A$  has a reason to  $\phi$  then  $A$  would, under certain specified circumstances, be motivated to (some degree to)  $\phi$ . Of course externalists who endorse this view will have a specific conception of what those circumstances are; and this conception is naturally going to be different from the one endorsed by internalists. As Susan Hurley puts it:

Those who endorse the view that  $R$  [('there is a reason for  $A$  to  $\phi$ ')] entails that  $M$  [('if  $A$  knew the relevant facts and was rational, she would be motivated to  $\phi$ ')] may do so for quite different reasons. Their underlying views about the relations between reasons and motivation may be very different. For example, someone could hold that  $M$  is true in virtue of  $R$ , a modification of the broadly Platonic view that you cannot truly know the good without loving it. Someone else could hold that not- $R$  is true in virtue of not- $M$ , a broadly Humean view that having a reason to act

consists of no more than some facts about actual or hypothetical motivation, so that the reason claim is false in virtue of the falsity of the relevant claims about motivation. Though both agree that *R* entails *M*, they do so for opposite reasons.<sup>1</sup>

Another reason why internalists and externalists may agree about *R*'s entailing *M* but for different reasons may stem from the fact that they have different conceptions of what is meant by 'rational deliberation'. In chapter one, I said that 'rational' was ambiguous and I distinguished between a formal and a substantive conception of rationality. This ambiguity, it seems, is here exported to the idea of rational deliberation. Discussing Williams (qua internalist), Brad Hooker says:

The external reasons theorist will...be dissatisfied with Williams' conception of rational practical deliberation - in contrast to Williams, the external theorist is likely to think that (at least some) rational deliberation about reasons for action starts not from the agent's own subjective present motivations, but from some objective ('external') values or requirements, fixed independently of the agent's present motivations. So, the external reasons theorist will deny that the content of the proposition that there is a reason

for one to  $\phi$  is to be captured by the proposition that, if one deliberated rationally according to Williams' conception of rational deliberation, one would be motivated to  $\phi$ .<sup>2</sup>

On the externalist conception of reasons (and rational deliberation), then, since an agent's reasons are not grounded in her motives, what then are they be grounded in? A natural, and not entirely implausible answer is that, as Hooker suggests, agents' reasons for actions are grounded in, derived from, or in some other way related to the goodness of the actions in question. A simple version of this theory would be: A has reason to  $\phi$  if and only if  $\phi$ -ing is (in some sense or other) good. Theories of this kind are usually called 'value-based' or 'teleological' theories of reasons. On this conception of reasons, it is *because*  $\phi$ -ing is good that we have a reason to  $\phi$ . A natural extension of this theory is the idea that the more valuable  $\phi$ -ing is the more reason we have to do it, and, consequently, what we have most reason to do is what is most valuable.

As we saw in chapter one, Prichard thought that it was a mistake to appeal to value-based reasons in an attempt to supply an answer to 'why should I be moral?' This belief was prompted by Prichard's scepticism about there being *any* kind of reasons we can appeal to in order to show that we ought to be moral (a scepticism which in turn stems from his intuitionism about moral obligation).

However, the view that reasons are value-based has also come under pressure in more recent discussions of the topic.

In *What We Owe to Each Other*, Scanlon argues that we should adopt a 'buck-passing' account of value. According to this account, reasons are not based on values; rather 'to call something valuable is to say that it has other properties that provides reasons for behaving in certain ways with regard to it.'<sup>3</sup> The sort of behaviour that is called for will vary from object to object (and thus from properties to properties); in some cases there will be reasons for us to admire the object, in others there will be reasons to respect it, to promote its existence and flourishing, to desire it and so on. As Scanlon rightly points out: 'Understanding the value of something is not just a matter of knowing *how valuable* it is, but rather a matter of knowing how to value it - knowing what kinds of actions and attitudes are called for.'<sup>4</sup>

However, this version of the buck-passing account should be contrasted with the one Scanlon presents a page later. He says:

...being good, or valuable, is not a property that itself provides a reason to respond to a thing in certain ways. Rather, *to be* good or valuable *is* to have other properties that constitute such reasons. ...this account...takes goodness and value to be non-natural properties, namely the purely formal, higher-order properties of having some



lower-order properties that provide reasons of the relevant kind.<sup>5</sup>

At first glance, these two accounts may appear to be the same, but in fact, they are very different. The first is an account of what we *mean* by the words 'good' and 'valuable'. The second version tells us what it *is* for something to be valuable. The buck passing account of value can therefore take two different forms:

*Analytically Reductive:* When we assert that x is valuable, what we mean is that x has other properties that provide us with a reason (or reasons) to adopt a pro attitude toward x.

*Non-analytically Reductive:* Though 'x is valuable' and 'x has properties that provide us with a reason (or reasons) to adopt a pro attitude toward x' do not mean the same, when 'x is valuable' is true that fact is the same as, or consists in, the fact that x has properties that provide us with a reason (or reasons) to adopt a pro attitude toward x.

It is unclear which account Scanlon takes himself to be providing. He says that he is led to the buck-passing account by reflections on the implications of Moore's open question argument about 'good'.

According to Moore, it is a mistake to try to provide a reductive analysis of the meaning of 'good'. His argument runs roughly as follows. Suppose some philosopher were to suggest that 'good' means 'conducive to happiness'. If these concepts were indeed synonymous, then the question 'x is conducive to happiness, but is x good?' would be equivalent to asking 'x is good, but is x good?'. But these two questions are not equivalent: the former is a perfectly legitimate question whereas the latter is, if not nonsensical, then at least very odd. But since this is so, it cannot be the case that 'conducive to happiness' means the same as 'good'. Importantly, the point can be generalised: for any proposed naturalistic<sup>6</sup> analysis of 'good', it will always make sense to ask whether an object (or an action, or a state of affairs) which instantiates this natural property really is good. Hence the meaning of 'good' cannot be analysed in terms of some natural property or properties. Moore of course thought that many things are good: happiness, friendship, and beauty are all good things according to Moore, but these things are good because they all have attached to them the unanalysable property of goodness.

However, Moore seems to have failed to distinguish the two kinds of analyses mentioned above in his discussion of 'good'. It

appears Moore thought he could derive substantive conclusions about the nature of goodness (namely, that it is *sui generis*) by analysing the meaning of the term 'good'. Many contemporary philosophers doubt that this can be done. To illustrate the difference between the two kinds of accounts consider the following well known example: water is  $H_2O$ , but 'water' does not mean ' $H_2O$ ' - 'water' means, roughly, 'the tasteless, odour-less, clear liquid that falls from the sky and can be found in lakes and rivers'. As it turns out this liquid happens to be  $H_2O$ . That water is  $H_2O$  is an empirical discovery, but this does not mean that people did not know what 'water' meant before it was discovered that water is  $H_2O$ .

Moore's 'Open Question Argument', if successful, shows at most that 'good' is not *analytically* reducible to anything else - i.e. it shows that the *meaning* of 'good' cannot be explicated in terms of some other concept or concepts. Even if this is correct, Moore has not demonstrated that the property of being good cannot be identical to, say, the property of being conducive to pleasure. It is therefore important to keep semantic and metaphysical analyses and/or reductions separate. The buck-passing account of value, conceived of as an account of the meaning of 'good' or 'valuable', does however put pressure on Moore's account of the 'semantic autonomy' of 'good' and 'valuable'. It says that the meaning of these words can indeed be captured by using the non-natural concept of 'a reason' (and the natural(?) concept of a 'pro

attitude').<sup>7</sup> Whether or not the buck-passing account succeeds in capturing the meaning of 'good' or 'valuable' I shall leave unexamined.

As a non-analytically reductive account of value, the buck-passing account may be put as follows:

For all  $x$ ,  $x$  is valuable if and only if  $x$  has other properties that provide us with a reason to adopt a pro attitude toward  $x$ .

It is of course true that  $x$  can be valuable in different ways:  $x$  can be instrumentally valuable - that is, valuable as a means to something else - or as an end in itself - that is, valuable for its own sake (or 'finally' valuable). Thus,  $x$  is valuable for its own sake iff  $x$  has other properties that provide us with a reason to adopt a pro attitude toward  $x$  for its own sake; and  $x$  is instrumentally valuable iff  $x$  has other properties that provide us with a reason to adopt a pro attitude toward  $x$  for the sake of its effects.

It seems uncontroversial that the left-to-right implication of the buck-passing account is correct: if  $x$  is valuable (either finally or instrumentally), then  $x$  has other properties that provide us with reasons to adopt a pro attitude toward  $x$ . However, it is not clear that the right-to-left implication is correct. There are philosophers who have argued that an object might be totally devoid of value

whilst still possessing properties in virtue of which we have a reason to have a pro attitude toward that object. Suppose some object *x* (a bottomless drinking glass say) lacks final value and that some eccentric millionaire has made a standing offer to anyone who forms a pro attitude toward *x* for its own sake that he or she will receive a million pounds for having this attitude. In this case it seems we all have a reason to form a pro attitude toward *x* for its own sake; yet *x* is, *ex hypothesi*, not finally valuable.<sup>8</sup>

Whether or not this is a good objection to the buck-passing account of value is for present purposes not very important. What is important, however, is the fact that Scanlon thinks that an object's being valuable is not *itself* a property of that object which gives us a reason to act in a certain way with regard to it. Although Scanlon does suggest that an object's being valuable can be analysed in terms of that object's having other properties such that they provide reasons for having a pro attitude toward that object, he is not, he says, providing an account of what it is for something to be valuable:

It is not a "theory" of value: neither a systematic account of which things are valuable, nor an explanation of the "source" of value.<sup>9</sup>

In light of this, I think we should interpret Scanlon's buck-passing account, not as a reductive account of value, but as a *redundancy theory* of value - i.e. as theory according to which reasons are not grounded in the value of an object (or action, or state of affairs); rather, reasons exist or obtain in virtue of the obtaining of those features of an object on which the value of that object supervenes. On this theory, the value of an object, event, state of affairs etc. is, we might say, *normatively epiphenomenal*. The idea is that if an action is good because it is kind, say, then it is the kindness of the action that provides us with a reason to approve of it and to encourage others to perform similar actions etc. It is not as if the kindness of the act makes it valuable and then that the value of the action gives us a reason to approve of it. Therefore it is false, or at least misleading, to say that reasons are *value-based*. It will of course be true on this theory that if x is valuable then there is reason to form a pro-attitude toward x, so it will be true on this theory that reasons are 'value-related'. I believe we should accept the buck-passing account of value understood as a doctrine about the redundancy (or normative epiphenomenalism) of value(s) with respect to reasons for action and reasons for forming pro attitudes etc. The only justification I can offer for my accepting the buck-passing account of value is that it strikes me as, simply, correct.<sup>10</sup>

So if reasons are not desire-based (as I tried to show in chapters two and three), and they are not value-based, then what

are they based on? The buck-passing account says that our reasons are based on, or stem from, or obtain in virtue of, the obtaining of those properties of an action on which that action's value supervenes. So what are these properties then? I have no general answer to this question, but I suspect no one else does either. I am not saying that such an answer cannot be found, but for my purposes in this dissertation I need not supply such an answer. Nonetheless, regardless of what the truth is about these reason-grounding properties, a serious problem for any attempt to answer the question 'why should I be moral?' still remains, and this problem certainly demands an answer. I will refer to this problem as 'Scanlon's dilemma'.<sup>11</sup>

### **4.3 Scanlon's dilemma**

Scanlon says that the question 'why should I be moral?' can be understood as a request for a demonstration that considerations of right and wrong are really reason-giving. However, in trying to show this, we face an awkward dilemma. As Scanlon puts it:

Attempts to explain how the fact that an action is wrong provides a reason not to do it face a difficult dilemma.

Understood in one way, the answer is obvious: the reason not to do the action is just that it is wrong. But this is



surely not the kind of answer that is wanted: it simply takes the reason-giving force of moral considerations for granted. Suppose, on the other hand, that we were to appeal to some clearly nonmoral reason, such as that people have reason to be morally good because, taking into account the effort that deception requires, the likelihood of being found out, and the costs of social ostracism, it is in their self-interest to be moral. This account might supply a reason for doing the right thing, but it would not be the kind of reason that we suppose a moral person first and foremost to be moved by.<sup>12</sup>

Although Scanlon's dilemma is difficult in its own right, there is an additional problem associated with attempts to answer 'why be moral?' which is not well brought out in the passage just quoted. Even if we can supply a reason for taking wrongness seriously, we will not have answered the moral question as I understand it. What we need to show in order to answer the moral question, is not merely that we have a reason to be moral (or to refrain from doing that which is wrong). Even if we can show this in a way which escapes Scanlon's dilemma, it doesn't follow that we *should* be moral. For all we know, the reason we have for being moral (if we do have such a reason) may be a very weak reason, a reason that

can, and perhaps often is, outweighed by reasons of self-interest if and when the two types of reasons conflict.

According to Scanlon, when we ask 'why be moral?' we might do so for at least two different reasons: the question may be asked with the aim of self-understanding, or with the aim of justification.<sup>13</sup>

He says:

The task of explaining how the fact that an action would be wrong provides a reason not to do it can be seen, first, as a task of self-understanding: we want to understand the reasons we are responding to when we are moved by moral considerations.<sup>14</sup>

But as he (in my mind rightly) points out:

...there seems to be more at stake than mere interpretation of the reasons we take ourselves to have. Even from the point of view of those of us who already care about right and wrong, a mere portrait of what it is we care about may seem to give us less than what we want: what we want to know is not merely what we care about when we care about right and wrong but why this is something we must care about. [...] This might be put saying that what the question "Why be moral?" calls for is

not mere self-understanding but justification: an account of why we and others have compelling reason to be moral.<sup>15</sup>

However, he says, 'justification' is a misleading term for what is needed here. It is misleading in two different ways. First,

[i]t is misleading to say that what those of us who already care about right and wrong are looking for in our own case is a justification, because this suggests that we think we should abandon our concern with right and wrong unless some additional ground for it can be provided.<sup>16</sup>

I must confess that I do not understand why it is misleading to ask for a justification for our concern about right and wrong. What does Scanlon mean by 'additional' ground here? Additional to what? Our *caring* about right and wrong? It is implausible to suggest that our caring about a certain institution or practice is in itself sufficient to legitimise or justify our continued concern with that institution or practice. Of course, it may be true of those who care about a particular practice that they are not concerned with providing a justification for it, but this has nothing to do with whether they should or should not give up that practice. Even if a self-proclaimed *Übermensch* felt no need to justify his cruel action to some

unfortunate person, it is still an open question whether he is justified (in the relevant sense) in engaging in acts of cruelty.

This brings us to Scanlon's second reason why he thinks calling this an aim of justification is misleading. Scanlon says:

It is also misleading to say that we are looking for a way of justifying the morality of right and wrong to someone who does not care about it - an "amoralist" - because this suggests that what we are looking for is an argument that begins from something to which such a person must be already committed and shows that anyone who accepts this starting point must recognize the authority of the morality of right and wrong.<sup>17</sup>

In one sense this is correct. If we can, through sound argument, justify morality to someone who is not an amoralist, why shouldn't this argument be equally sound for an amoralist? If we can produce a rationally compelling argument which shows that the normative force of moral considerations is such that it trumps all other considerations, why should it matter if an amoralist is left cold by our argument? If justifying morality to an amoralist consists in providing him with a psychologically compelling answer (like Prichard mistakenly thought), the prospect of finding such a justification may be very bleak indeed. On the other hand, why

should we think that justifying morality to an amoralist must proceed via something he already cares about? This would be the case, I think, only if some version of internalism about reasons was roughly right, but fortunately this is not so.

Since he rightfully doubts that a justification of the former kind can always be provided, what should be sufficient to allay our 'reasonable concerns', Scanlon says, 'is a fuller explanation of the reasons for action that moral conclusions supply.' However, 'In giving this explanation [...] we must address the problem of the moral "must" - the seeming necessity of moral concerns....'<sup>18</sup>

Explaining this 'must' will then consist in explaining why considerations of wrongness always (or almost always) takes priority over other values and why failing 'to see the reason-giving force of such considerations strikes us a particularly serious fault.'<sup>19</sup>

But if an explanation of the moral 'must' is a part of what we are looking for when our aim is self-understanding, a natural question to ask is how the aim of self-understanding differs from the aim of justification. If we aim to 'understand the reasons we are responding to when we are moved by moral considerations', we are of course presupposing that there are such things as moral considerations (how else could we be moved by them?) and that these considerations provide us with genuine reasons (if they did not then the aim of understanding them would be an exercise in futility). A sceptic could of course deny that are any moral

considerations; or, alternatively, he could hold that although there are such things as moral considerations, these considerations do not provide us with reasons for action. But if we presuppose that there are moral considerations and that these considerations provide us with reasons to act, what is left of the aim of self-understanding? 'To explain why the moral "must" always takes priority over other values' seems to be the obvious answer. However, if we really can explain why moral considerations take precedence over other considerations (in a way that escapes Scanlon's dilemma) we will, *ipso facto*, have supplied a justification for morality. What else could a justification for morality consist in?

#### **4.4 Scanlon's contractualism as an answer to the dilemma**

Before we can assess Scanlon's answer to the moral question we first need to know what Scanlon's conception of wrongness is. His contractualist formula (SC) runs as follows:

An act is wrong if and only if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behavior that no one could reasonably reject (*from an individual standpoint*) as a basis for informed, unforced general agreement.<sup>20</sup>

For the sake of exposition, I shall call the property of being 'disallowed by any set of principles for the general regulation of behavior that no one could reasonably reject (*from an individual standpoint*) as a basis for informed, unforced general agreement' (RR).

So what is this formula supposed to show? On one interpretation it is intended to reveal (merely) the *nature* of wrongness - i.e. what it *is* for an act to be wrong. On another interpretation, (SC) purports to show what the *ground* of wrongness is - i.e. what *makes* an act wrong.<sup>21</sup> To see the difference between these interpretations, consider the case of what it is for something to be fragile. An object, *x*, is fragile if and only if *x* would break in certain circumstances - this is *what it is* for something to be fragile. However, the grounds for *x*'s fragility - i.e. what *makes* *x* fragile - is *x*'s micro-physical structure.

Scanlon vacillates between these interpretations of (SC). Sometimes he seems to suggest that (SC) should be given a stronger reading according to which it reveals the ground of wrongness. He writes:

a defence of contractualism has to argue that the idea of justifiability to others can be seen to play an important role in shaping our thinking about right and wrong, and that particular moral arguments seem to establish that an



action is wrong just when, *and just because*, they show that so acting could not be justified to others on grounds they could not reasonably reject.<sup>22</sup>

According to this version of (SC), an act is wrong *because* it has property (RR). There are as far as I am aware at least two objections to this interpretation of (SC). We may call these the *Euthyphro* objection and the *Redundancy* objection, respectively. The first kind of objection is voiced by Judith Thomson who writes

I cannot bring myself to believe that what *makes* it wrong to torture babies to death for fun (for example) is that doing this 'would be disallowed by any system of rules as a basis for the general regulation of behavior which no one could reasonably reject as a basis for informed, unforced general agreement.' My impression is that the explanation goes in the opposite direction - that it is the patent wrongfulness of the conduct that explains why there would be general agreement to disallow it.<sup>23</sup>

It is natural to think of objections of this sort as being motivated by a kind of intuitionism about wrongness. According to this kind of intuitionism, wrongness is a simple, non-natural, unanalysable property similar to the property of goodness that Moore thought he

had identified. Scanlon obviously - and for good reason - wants to avoid this kind of intuitionism. Apart from epistemological worries about how we come to know about these properties, since one of Scanlon's aims is to 'characterize wrongness in a way that makes clear what reasons wrongness provides'<sup>24</sup> it is obvious why he can't accept that wrongness is unanalysable. I shall, however, argue that Scanlon's contractualism cannot provide such a characterisation. To see why this is so we need to take a look at the redundancy objection to (SC).

If an act is wrong *because* it has property (RR), then obviously we cannot reasonably reject a principle on the grounds that (or because) the action it prescribes is wrong. But this is where the charge of redundancy comes in. As Michael Ridge puts it:

whenever principles allowing an action are reasonably rejectable because such actions have feature F, such actions are wrong simply in virtue of having F and not because their having F makes principles allowing them reasonably rejectable.<sup>25</sup>

So if we can reasonably reject a principle on the grounds that the act it prescribes is unfair, say, then it seems that it is the unfairness of the act - and not its having the property (RR) - which makes it

wrong. But if this is so, an act's having the property (RR) is redundant with respect to that act's wrongness.

Furthermore, even if having property (RR) should be understood as specifying the grounds of wrongness, then wrongness must be a separate property from property (RR), and the reason-giving force of wrongness *itself* will not have been explained. In response to these objections, Scanlon says (inconsistently with what he says on p. 11 of *What We Owe to Each Other*):

The contractualist formula...is intended as an account of what it is for an act to be wrong. What *makes* an act wrong are the properties that would make any principle that allowed it one that it would be reasonable to reject.<sup>26</sup>

Here Scanlon says that his contractualism should be understood as merely specifying the necessary and sufficient conditions (i.e. the nature of wrongness) for an act's being wrong which is *not* the same as specifying the ground of wrongness. On this interpretation, Scanlon's contractualism is not open to the two objections just mentioned (or at least not obviously so) since those objections arise only if (SC) is understood as aiming to specify the ground of wrongness.

However, even if Scanlon's contractualism should be understood merely as an account of the nature of wrongness (as opposed to the ground of wrongness), by holding onto his intuition that 'the fact that an act is wrong seems *itself* to provide us with a reason not to do it'<sup>27</sup> his account is still vulnerable to a version of the redundancy objection. Suppose (for reductio) that an act's having the property of being wrong provides us with a reason not to do the thing in question. If the property of being wrong is *identical* to property (RR), then an action's having property (RR) must provide us with a reason not to perform that action. But it does not seem to. If an action has property (RR) because it has feature F, then the reason we have for not doing the thing in question is provided by the fact that the action has feature F, not that it has property (RR). Therefore, having property (RR) seems redundant with respect to the reasons we have for not doing the thing in question. So if having property (RR) is identical with being wrong, then since having property (RR) is not reason-giving, an act's being wrong is not reason-giving either. Hence Scanlon cannot explain how an act's wrongness provides us with a reason not to do it. And the reason why he cannot explain this is (as I am going to argue) that an act's wrongness *does not* provide us with a reason not to do the thing in question.

It is somewhat surprising that although he endorses a buck-passing account of value, Scanlon explicitly rejects a buck-passing

account of wrongness. Unlike the property of being valuable, apparently, the property of being wrong is in itself reason-giving according to Scanlon. However, as Stratton-Lake has recently pointed out<sup>28</sup>, if he abandons the view that wrongness is in itself reason-giving, he can escape both the *Euthyphro* and the *Redundancy* objection. In other words, Scanlon should hold the view that (SC) only tells us what the nature of wrongness is (thereby avoiding both the *Euthyphro* and *Redundancy* objections) and adopt a buck-passing account of wrongness (thereby giving himself a chance, at least, to explain why considerations that make acts wrong also give us reason not to do wrong acts). He could reformulate his aim so that it was no longer to characterise wrongness in a way that makes clear what reasons wrongness provides, but rather to show why considerations in virtue of which actions are wrong provide us with reasons.

Scanlon himself hints at what such an account of wrongness would look like. He says that on such an account wrongness would be 'the higher order property of violating (some or other) important standards of conduct and therefore being open to (some or other) serious objection.'<sup>29</sup> However, even if Scanlon abandons his original intuition and adopts a buck-passing account of wrongness, he still needs to explain how contractualism can account for the priority of moral reasons over other kinds of reasons. In other words, he still needs to explain why we *should* (or *ought* to) be moral.

In the remainder of this chapter I will try to provide at least a preliminary sketch of what I think a successful buck-passing account of wrongness will look like. I think a such an account can supply the right kind of answer to the moral question. It can do this in a way which escapes both the dilemma that Prichard discussed (i.e. the dilemma which arises if we attempt to give a purely teleological justification/explanation of why we should be moral) and the dilemma that Scanlon attributes to Prichard. The account will be consistent with the metanormative assumptions I made in chapter one and it will show that acting morally is rational in the strongly substantive sense of 'rational' identified in chapter one; i.e. we are, in fact, fully practically justified in acting morally. Although my account shares some features of Scanlon's account, it is distinct from it. Unfortunately, I cannot provide an iron-clad defence of my account in the remainder of this dissertation - some questions will have to be left unanswered. Nonetheless, I think the particular kind of buck-passing account of wrongness I am going to sketch goes a long way toward answering the moral question.

#### **4.5 A buck-passing account of wrongness?**

As we have seen, buck-passers about value believe that the property of being valuable is not itself a reason-giving property. Rather, to be valuable is to have other properties that provide us

with reasons to respond to valuable things in a certain way. So what are the merits of a buck-passing account of wrongness? Consider the following *generic* non-analytically reductive buck-passing account of wrongness (an account which is strictly analogous to the non-analytically reductive buck-passing account of value):

...being wrong, is not a property that itself provides a reason to respond to an action in certain ways. Rather, for an act to be wrong is for it to have other properties that constitute such reasons. ...this account...takes wrongness to be a non-natural property, namely the purely formal, higher-order property of having some lower-order properties that provide reasons of the relevant kind.

I call this a 'generic' account because on the face of it, this account seems 'formal' enough to be acceptable to a wide range of moral theorists: Kantians, utilitarians, and virtue ethicists should all be able to accept this account. These theorists will of course disagree about what the relevant lower-order properties are, but this does not mean that they couldn't all accept this generic version of the buck-passing account of wrongness.

If we accept a buck-passing account of wrongness we have a way of avoiding Scanlon's dilemma. Recall that the dilemma was



supposed to show that all attempts to explain why an act's wrongness gives us reason not to do it are either circular or of the wrong kind. But now we should be able to see that this is not a dilemma for the buck-passing account of wrongness. This account *does not* try to explain why the fact that an action is wrong provides a reason not to do it. This is precisely what the account denies and what 'passing the buck' means: the wrongness of an act *does not* provide us with reason not to do the act; rather it is those considerations in virtue of which an act is wrong which provide us with reason not to do the act in question. On this account, wrongness is (like value, or being valuable) normatively epiphenomenal: the fact that an act is wrong merely 'signals' that this act has properties which provide us with reason not to do it.<sup>30</sup>

But does this not merely postpone the problem? On the one hand, the relevant lower-order property of a wrong action in virtue of which we have a reason not to perform such an action cannot be the property of being wrong - if it were, the buck-passing account of wrongness would either be viciously circular or not a buck-passing account at all. On the other hand, if the lower-order properties of wrong actions (whichever ones they are) are such that they give us only some clearly non-moral reason not to perform such actions, these properties seem to provide us with the wrong kind of reasons - how could an act's wrongness consist in its having

a property which gives us a non-moral reason for not performing such an act?

Fortunately for the buck-passer of wrongness, however, although the relevant lower-order reason-giving property of wrong acts can neither be the property of being wrong, nor a property such that it gives us a non-moral reason for performing acts that are wrong, these two kinds of properties do *not* exhaust the logical space of properties (unlike moral and non-moral reasons which jointly *do* exhaust the logical space of reasons). There can still be properties - other than the property of being wrong - which provide us with reasons of the requisite kind. After all, there are other moral properties besides the property of being wrong.

So what are these properties then? My contention is that the relevant properties are those properties that can be accurately described by using 'thick' ethical concepts like 'cruel', 'inconsiderate', 'unfair' etc.. About such concepts, Williams says:

[Thick ethical concepts] are characteristically related to reasons for action. If a concept of this kind applies, this often provides someone with a reason for action, though that reason need not be a decisive one and may be outweighed by other reasons.... We may say, summarily, that such concepts are "action-guiding."<sup>31</sup>

Two things are of importance in this passage. First, Williams says that 'if a concept of this kind applies [to an action], this *often* provides someone with a reason for action.' This may be taken to express scepticism about the categoricity (or universality) of thick ethical concepts (and, by implication, of morality). Second, Williams also claims that the reason provided by considerations accurately describable by using thick ethical concepts need not be a decisive one and may be outweighed by other reasons. If Williams is right about this, the buck-passing account of wrongness does not seem like it can provide a plausible answer to the moral question. So these are obviously important challenges, and I shall address them both in due course.

For now, we should note that insofar as we think that some actions can accurately be described by using these concepts, then on the account under consideration, the lower-order reason-giving property of an action in virtue of which it is wrong, will not be the property of being wrong, so the account will not be circular. On the other hand, the reasons which properties accurately describable by thick ethical concepts provide us with *do* seem to be reasons of the right kind; they provide us with reasons we can easily see to be closely connected with what it means for an act to be wrong. On this picture, the buck-passing account of wrongness would run as follows:

An act's wrongness consists in the purely formal, higher-order property of having some lower-order property or properties - accurately expressible by, or accurately referred to by, thick ethical concepts - that give us reason for refraining from performing such actions.

An act would be wrong, then, because it is cruel, or inconsiderate, or unfair etc.

Crucially, however, this could be so only if considerations of cruelty, inconsiderateness, and unfairness etc. are genuinely reason-giving. What reason do we have to believe that these kinds of considerations actually provide us with reasons not to do the things in question? We can begin to answer this question by asking, somewhat defensively, 'what reason do we have to believe that such considerations are *not* reason-giving?'. If we dismiss the internalist answer to this question (as I have already argued that we should), then apart from answers stemming from a general scepticism about reasons and 'the normative' - which, in chapter one, were shown to be at best implausible and at worst simply incoherent - I can't think of any good reasons for denying that considerations of cruelty, inconsiderateness, and unfairness etc. are genuinely reason-giving. So much for the defensive answer. Is there anything more positive to be said about the thesis under consideration?

Well, common sense and everyday usage of these concepts certainly seem to support the idea that considerations of cruelty, inconsiderateness, and unfairness are reason-giving. From a philosophical point of view, however, this may seem inadequate; what else may be said in support of this thesis? I believe that propositions like

The fact that  $\phi$ -ing is cruel gives us reason not to  $\phi$

The fact that  $\phi$ -ing is unfair gives us reason not to  $\phi$

The fact that  $\phi$ -ing is inconsiderate gives us reason not to  $\phi$

are *conceptual* truths: if B were to say 'A's  $\phi$ -ing was cruel, but there was no reason for A not to  $\phi$ ', the correct reaction to B's utterance is not that B has made a substantive mistake about A's reasons; rather, the correct judgement would be that B is either insincere or not a competent speaker of English: anyone who understands the meaning of 'cruel' also thereby understands that there is reason not to perform actions accurately described as such.

It is of course true that ' $\phi$ -ing is cruel' is not *synonymous* with 'there is reason not to  $\phi$ ' - after all we may have reason not to  $\phi$  because of other considerations besides cruelty - but this lack of synonymy is no basis for criticism of this version of the buck-passing account of wrongness. For the purposes of offering such an account it is sufficient to show only that 'there is reason not to  $\phi$ ' is

*part of the meaning of 'φ-ing is cruel'.* 'Unmarried' is not synonymous with 'bachelor', but 'unmarried' is nonetheless part of the meaning of 'bachelor'. On the account under consideration, the same is true for 'there is reason not to' and 'is cruel' - at least in the sense of 'cruel' that we usually have in mind when describing certain actions (or persons).

As I have just said, those of us who endorse this view must of course hold that internalists (who would say that *'the fact that φ-ing is cruel gives us reason not to φ'* is not true) are somehow conceptually confused: they do not understand 'cruel' in the same way that ordinary, mature, competent speakers of English understand that word. Williams, however, suggests that there may be a different explanation (other than linguistic incompetence or conceptual confusion) for why actions accurately describable by thick ethical concepts can fail to provide agents with reasons of the relevant kind. He says:

It may well be that 'thick' ethical concepts are, to an adequate degree, both 'world-guided' and 'action-guiding'. People who use a given concept of this sort will find their application of it guided by their experience, and also accept that it gives them reason for and against various kinds of action. [...] But this does not mean that a speaker who does use a given concept of this kind...can truly say that

another agent who does not use the concept has a reason to avoid or pursue certain courses of action in virtue of that concept's application. To show this, the speaker would need to show that the agent *has reason to use that concept*, to structure his or her experience in those terms.<sup>32</sup>

This comment helps us understand what Williams meant by the comment that thick ethical concepts are *often*, but presumably not always, reason-giving: they are reason-giving for those, and only those who choose to use them, or for those who choose to structure their experiences in those terms. We should of course expect Williams to say this given his internalism about reasons, but why should we accept it? This appears to be nothing but internalist dogma. Whether or not a given concept *applies* to a particular action does not depend on whether an agent who performs the act in question 'has reason to use' that concept.

It may be that, for some odd reason, an agent has good reason not to *use* a particular concept, or good reason not to structure his or her experience by *using* such concepts, but this has nothing to do with whether or not the concept *applies* to the agent's action. Words, and the concepts they express, do not cease to mean what they mean just because some individual chooses not to use them (or because he has good reason not to use them). If an



action has the 'world-guided' and 'action-guiding' properties required for that action's being correctly described as cruel (whatever they may be), then this will be so regardless of whether a particular agent chooses not to describe the action in those terms. I see no reason why we should accept the claim that thick ethical concepts apply (in the requisite sense) only to those who have reason to use, or structure their experiences in terms of those concepts. In other words, I cannot see that the categoricity of the normative force inherent in thick ethical concepts (or, more precisely, the property or properties they refer to) is threatened or undermined by Williams' comments.

It should of course be admitted that we may have to allow for the possibility that some speakers of English may be 'conceptually impoverished'; e.g. young children and mentally handicapped persons may not know the meaning of some thick ethical terms. This may lead to some complications for the account under consideration, but I cannot consider this issue here as it would take us too far afield from our present concerns.

#### **4.6 Scanlon's dilemma revisited**

Scanlon believes that his conception of wrongness and the reasons that (as he sees it) flow from it can explain why we should be moral in a way which is connected closely enough with our idea of right

and wrong to avoid the 'wrong kind of reason' horn of the dilemma. At the same time, he says, the reason we have for wanting to stand in a 'justifiability relation' to others is not so closely identified with moral reasons that it impales itself on the circularity horn of the dilemma.<sup>33</sup> To illustrate Scanlon's position it will be useful to quote him at length. He says:

Instead of asking "Why be moral?" we might ask "Why be loyal to one's friends when this requires sacrificing other goods?" Considering the answer to this question will help cast light on the general problem of moral motivation. It may seem that in answering the question "Why be loyal?" we face an analogue of [Scanlon's dilemma]. The answer "Because friendship requires it." seems to be no response at all to the question that is being asked. But if, on the other hand, we cite some value other than friendship - if, for example, we appeal to the benefits of having friends - then this seems the wrong kind of response. A person who was "loyal" for that kind of reason would not be a good friend at all.

The right response to this dilemma is, first, to characterize the relationship that friendship involves in a way that makes it clear why it is something desirable and admirable in itself. Given such a characterization, we can

then see how, on the one hand, being a friend will also bring other benefits (such as enjoyable companionship, help, and support) and why, on the other, being a friend involves seeing "because loyalty requires it" as a sufficient reason for doing something even though it involves a sacrifice of other goods. By bringing these two elements together as aspects of a single value, such an account enables us to see that the analogue of [Scanlon's dilemma] is not really a dilemma at all. It merely appears to be one because it presents two essential aspects of friendship as if they were competing answer to the same question.<sup>34</sup>

So, in the case of being moral:

The contractualist ideal of acting in accord with principles that others (similarly motivated) could not reasonably reject is meant to characterize the relation with others the value and appeal of which underlies our reasons to do what morality requires. This relation...might be called a relation of mutual recognition. Standing in this relation to others is appealing in itself - worth seeking for its own sake. A moral person will refrain from lying to others, cheating, harming, or exploiting them, "because these things are wrong". But for such a person these requirements are not

just formal imperatives; they are aspects of the positive value of a way of living with others.<sup>35</sup>

I think we should reject Scanlon's 'explanation' of why we should be moral. Consider first his analogy with friendship. Although Scanlon is absolutely right when he says that it is part and parcel of being a (genuine) friend that one sees friendship as having both these 'pushing' and 'pulling' aspects. But this seems to get us less than what we want. The case of friendship shows only that friends *see* (or *treat*) considerations of friendship as having these two aspects. But this is, at best, only an observation about the 'psychology of friendship'. Just because friends *see* considerations stemming from friendship as being reason-giving, this does not mean that such considerations *are*, in themselves, reason-giving.<sup>36</sup> The same, it seems, is true of the moral person. Of course a moral person will see moral considerations as being reason-giving - this is part and parcel of what it means to be *moral*. This person may also, for all I know, enjoy the benefits of being a member of the moral community, but this seems, at best, to be a desirable *consequence* of being moral - these benefits do not seem to form the normative basis for being a moral person. Nothing Scanlon has said so far demonstrates that moral considerations *are* genuinely reason-giving.

On Scanlon's proposal, there are two possibilities. On the one hand, if the reasons moral agents take themselves to be responding to when refraining from doing wrong actions are genuine reasons, then appealing to the value of standing in a justifiability relation to others is redundant with respect to answering the question 'why should I be moral?'. People should be moral because being moral consists in, or at least involves, seeing moral considerations in their true light - i.e. seeing them as being reason-giving.<sup>37</sup> The admittedly desirable state of standing in a justifiability relation to others does not 'underlie' our reasons to be moral; rather, our standing in this relation to others is, again, a consequence of our (and others') acting morally.

On the other hand, if the considerations moral agents take themselves to be responding to when refraining from doing wrong actions are *not* genuinely reason-giving, it is hard to see how an appeal to the benefits associated with standing in a justifiability relation to others could explain why we should be *moral*. It seems reasonable to suppose that we can enjoy the benefits of standing in this relation to others if and only if other agents *believe* that we stand in this relation to them. But if this is so, and since we could presumably secure the same benefits by (deceitfully) convincing others that we stand in this relation to them, it is doubtful whether it is our *actually* standing in this relation to others that is the source of these reasons. As Scanlon himself is well aware of, someone

who cheated his way into the moral community could not by any stretch of the imagination be called *moral*. Considerations of actual or possible benefits simply do not provide the right sorts of reasons. So Scanlon's appeal to the dual aspects of being justifiably related to others is either redundant or of the wrong kind - or both!<sup>38</sup>

However, with the buck-passing account of wrongness in place, we don't need to appeal to any reason (moral or otherwise) we may have for wanting to stand in a justifiability relation to others in order to explain why we should be moral. As I said in chapter one, a moral person is someone who structures her life in accordance with the demands of morality; i.e. she is a person who, at least as long as she is rational (i.e. under normal circumstances), refrains from doing what she believes to be wrong. Now, although it is true that the moral person is someone who (at least normally) refuses to do that which is wrong, a morally *enlightened* person (i.e. someone who has come to accept the buck-passing account of wrongness) will refrain from lying to others, cheating, harming, or exploiting them, *not because these things are wrong*, but because (being a competent speaker of English) she can see that she has reason not to perform actions that are accurately described as deceitful, harmful, or exploitative etc.

Furthermore, and in contrast with Scanlon's account of the motives of the moral person, the buck-passing account is consistent with (and may even lend support to) the idea that there is

something objectionable about a person who refrains from cheating, lying, and stealing *simply* because he believes these things are *wrong*. Someone whose moral motivation was triggered solely by considerations of right and wrong could be accused of suffering from a kind of fetishism about wrongness.<sup>39</sup> When an agent judges that some act is wrong, she usually (always?) has some idea about why, or in virtue of what, this act is wrong. In normal cases, such an agent will be motivated not by the judgement that the act is wrong, but by those considerations in virtue of which she takes the act to be wrong. Scanlon appears to disagree. He says:

When, for example, I first read Peter Singer's famous article on famine and felt the condemning force of his arguments, what I was moved by was not just the sense of how bad it was that people were starving in Bangladesh. What I felt, overwhelmingly, was the quite different sense that it was *wrong* for me not to aid them, given how easily I could do so. It is the particular reason-giving force of moral wrongness that we need to account for.<sup>40</sup>

This, I believe, is not a correct description of most people's natural reaction to the plight of the starving people in Bangladesh. I think that the natural and spontaneous reaction to cases like these is that it would be selfish or inconsiderate, or inhumane not to help these



people, given how badly off they are and how easily we could help them. This reaction can, and often does, illicit the judgement that it would be wrong for us not to help these people. But this judgment is merely a consequence of the spontaneous (and, I would add, warranted) judgement that it would be inhumane not to help these people. It seems to me that an agent who, whilst acknowledging that  $\phi$ -ing is inconsiderate, would not be prepared to refrain from  $\phi$ -ing until she had convinced herself that  $\phi$ -ing is wrong, would, in Williams' famous words, be having 'one thought too many'.<sup>41</sup>

However, even with the buck-passing account in place, we have not yet established that we *ought* to be moral. To do so we must show that Williams' second (sceptical) proviso in his discussion of thick ethical concepts is incorrect. This is the proviso that the reason provided by considerations accurately expressed by thick 'ethical concepts' 'need not be a decisive one and may be outweighed by other reasons'. In other words, the buck-passer about wrongness, it seems, still has to explain the moral 'must'.

#### **4.7 'Reason', 'a reason', and 'ought'**

Consider the following version of the buck-passing account of wrongness:

An action's wrongness consists in the purely formal, higher-order property of having some lower-order property or properties - accurately expressible by, or accurately referred to by, thick ethical concepts - which is such that an action's having this property *makes it the case that we ought not* perform such actions.

This seems to be what we need. What can be said in defence of this account? Before I can answer this question I must say something about the relation between the concepts 'a reason', 'reason', and 'ought'. On the version of the buck-passing account we have been working with in the last few sections, the key normative concept is that of 'reason'. I have said things like 'the fact that  $\phi$ -ing is cruel gives us *reason* not to  $\phi$ ' but I have not said much about the normative concept 'reason' itself. It is important to distinguish 'reason' from 'a reason'. I have used (or at least intended to use) 'reason' as a generic normative concept. Used in this way, when we say that some consideration gives us reason to  $\phi$ , all this means is that some unspecified ('positive') normative relation holds between this consideration, us, and  $\phi$ -ing. This unspecified normative relation can be an 'ought' or a pro tanto reason (a normative entity the nature of which we will turn to presently) or anything else which is normative. I think that a proposition like 'the fact that Bob would enjoy going to the match

gives Bob a (pro tanto) reason to go to the match' implies 'the fact that Bob would enjoy going to the match gives Bob reason to go to the match' but the opposite implications do not hold. There can be cases where it is true that there is reason for us to do something without its being the case that there is a pro tanto reason for us to do the thing in question. In this section I shall explain why I think this is so.

I said in chapter one that although all satisfactory answers to questions of the form 'why ought I...?' must be sought in a sound theory of practical reason, it is far from obvious that a sound theory of practical reason is reducible to a sound theory of practical reasons - i.e. to a sound theory of pro tanto reasons. In other words, it is far from obvious that, as Raz puts it, 'The normativity of all that is normative consists in the way it is, or provides, or is otherwise related to reasons.'<sup>42</sup> When Raz speaks of 'reasons' in the plural it is natural to interpret him as thinking of these as being pro tanto reasons. If this is what Raz is saying, then he is mistaken. Questions of the form 'why ought I...?' are equivalent to questions of the form 'in virtue of what, if anything, is it true that I ought to...?'. It is easy to see why philosophers (like Scanlon) who share Raz's convictions concerning the centrality and importance of the concept of a reason should think that 'ought' statements are made true (when they are true) by there being more reasons favouring the action in question than there are reasons favouring

not performing it. This is a natural assumption to make if you believe that, as John Broome puts it, 'All is reasons.'<sup>43</sup> Broome, however, does not share this belief. He says:

[Pro tanto] Reasons are undoubtedly important, but normativity has other important features, and our preoccupation with [pro tanto] reasons distracts us from them.<sup>44</sup>

I think Broome is right about this. Broome says that 'The key to understand the concept of a reason is to look at how facts of a particular type are explained'.<sup>45</sup> The facts in questions are what he calls 'ought facts'.<sup>46</sup>

Now, as Broome points out, much of the confusion surrounding the concept of a reason stems from a confusing feature of English which involves the concepts 'reason' and 'explanation'. When the fact that X explains why it is the case that Y, we often say that X is the reason for Y. In 'X is the reason for Y', however, 'reason' is used non-normatively. If we say that the reason why the house burnt down was that it had faulty electrical wiring, we are not saying that house's having faulty electrical wiring was a normative reason for it to burn down. This would be silly: houses don't have normative reasons - only agents do. Thus when we say 'the reason why you should be moral is X' we are simply saying that

you ought to be moral and X explains why this is so. This is an unwelcome feature of English, but since, as Broome points out, we very often cannot resist the slide from the non-normative sense of 'reason' to the normative sense of 'reason' (thus creating a normative sense of 'is the reason for you to  $\phi$ '), we must take this normative sense of 'reason' into account and clearly mark off its boundaries. Broome does this by calling reasons of this kind 'perfect reasons'.<sup>47</sup> Thus 'X is a perfect reason for you to  $\phi$ ' should be understood as 'you ought to  $\phi$  and X is the explanation of why you ought to  $\phi$ '.

However, we sometimes say that X is a reason for you to  $\phi$  when it is not the case that you ought to  $\phi$ . In this case X does obviously not explain why you ought to  $\phi$  since it is, *ex hypothesi*, not the case that you ought to  $\phi$ . The kind of reason we have in mind when we say this, then, cannot be a perfect reason - instead, it is a pro tanto reason. In chapter one I discussed some ways in which pro tanto reasons can be weighed against each other. For Broome, the weighing dimension of pro tanto reasons is essential to them. To explain this feature of pro tanto reasons, Broome uses an analogy with mechanical weighing: just like the weights in the pans of a pair of scales determine how the scales tip, so too do the weights of pro tanto reasons determine what we ought to do. Suppose we are wondering whether we ought to  $\phi$ . Sticking with the analogy, we should be able to determine what to do by

watching how the pro tanto reasons in favour of  $\phi$ -ing weigh against the pro tanto reason in favour of not- $\phi$ -ing. What could happen is that

The numbers associated with the reasons to  $\phi$  add up to more than the numbers associated with the reasons not to  $\phi$ . That is why you ought to  $\phi$ . [...] When an explanation of why you ought to  $\phi$  takes this form, I shall call it a 'weighing explanation'. The idea of a pro tanto reason arises in the context of a weighing explanation.<sup>48</sup>

The idea, then, is that a pro tanto reason is a 'functional' entity; it is something which plays a particular role in a weighing explanation of ought facts. More specifically, a pro tanto reason for someone to  $\phi$  is an entity which, as Broome puts it, plays the 'for- $\phi$ ' role in a weighing explanation of an ought fact. Importantly, since we can have reason to  $\phi$  even though it is not the case we ought to  $\phi$ , the reason we have to  $\phi$  must play the for- $\phi$  role in a weighing explanation of either (I) why we ought to  $\phi$ , or (II) why we ought not to  $\phi$ , or (III) why it is not the case that we ought to  $\phi$  and it is not the case that we ought not to  $\phi$ . On this conception of reasons

Any weighing explanation of why you ought to  $\Phi$  must include at least one pro tanto reason for you to  $\Phi$ . This reason must have a weight. But sometimes the explanation will include no other reason - in particular no reason for you not to  $\Phi$ . In that case, you ought to  $\Phi$ , whatever the weight of the reason for you to  $\Phi$ . But if that is so, how does the weight of this reason play any role in the explanation? It plays a role counterfactually. If there *were* other reasons for you to  $\Phi$  or not to  $\Phi$ , and this reason still held and had the same weight, its weight would participate in determining whether or not you ought to  $\Phi$ . This would only be so if, were the contrary reasons weighty enough, it would be the case that you ought not to  $\Phi$ . So a characteristic of any pro tanto reason is that it is possible for it to be outweighed. Indeed, this is what the term 'pro tanto' implies.<sup>49</sup>

Broome's theory of reasons is unorthodox in that it denies the centrality and importance of (pro tanto) reasons to practical reason. We have already seen that authors like Scanlon, and Raz hold that the concept of a reason is primitive - that it cannot be analysed in terms of other normative (or non-normative) concepts. As you may recall, Scanlon, for instance, says



I will take the idea of a reason as primitive. Any attempt to explain what it is to be a reason for something seems to be to lead back to the same idea: a consideration that counts in favour of it. "Counts in favour how?" one might ask. "By providing a reason for it" seems to be the only answer.<sup>50</sup>

Broome, on the other hand, takes the concepts of 'ought' and 'explanation' to be primitive. Rather than showing that Scanlon et. al. are wrong, I want to offer Broome's theory as an alternative to the orthodox view (presumably endorsed by Scanlon and Raz).

Broome himself has surprisingly little to say about why we should accept his 'functionalist' account of reasons over the 'pro tanto reason-as-primitive' view. He says that the preoccupation with reasons has led some authors to confuse 'wide scope' oughts with 'narrow scope' reasons (as in the case of his 'normative requirements'),<sup>51</sup> but although he is probably right about this, we certainly cannot, on this basis, draw the conclusion that the pro tanto-reason-as-primitive view is false. That would be an unacceptable *ad hominem* way of arguing for Broome's theory. The closest Broome comes to actually producing an *argument* for his functionalist view (in opposition to the reason-as-primitive view) is in the following passage:

'The reasons for action are considerations which count in favour of that action'. [...] These are Raz's words, but the remark is commonplace. [...] My definition of a pro tanto reason is a version of it, tightened up enough to make it a genuine definition. The definition specifies just what counting in favour of amounts to. To count in favour of  $\phi$  is to play a particular role in an explanation of why you ought to  $\phi$ . Counting in favour of is sometimes thought to be the basic normative notion. [...] But it cannot be, because it is complex. It incorporates the two elements of normativity and explanation. The notion of a reason has the same complexity.<sup>52</sup>

This may all be true, but it cannot count as an argument *in favour* of Broome's position. To say that the two notions of 'counting in favour of' and 'a reason' are complex - that they incorporate the two elements of normativity and explanation - is, in the present context, question begging against Scanlon et. al.; it simply presupposes the truth of functionalism.

In my view, there are very few, if any, arguments that can be brought to bear on the question whether pro tanto reasons or oughts and explanation are primitive. I accept functionalism for two reasons. First, Broome's own example of a 'normative requirement' does seem to be genuinely normative but the

normativity of this requirement does not seem to be reducible to (some function of) pro tanto reasons. Second, it simply doesn't seem right that figuring out what one ought to do is simply (and always) a matter of weighing up the pros and cons of doing this thing. When I reflect on the fact that it would be cruel for me to intentionally humiliate someone in public for the sake of eliciting a cheap laugh, this consideration and its normative force, it seems to me, could not be outweighed by other considerations. When we deliberate about cases like these, as soon as the thought that an action would be cruel enters our deliberation, this usually settles the matter. That the act would be cruel seems sufficient in its own right to determine what we ought not do it. In the next section, I shall try to argue that this is indeed so.

#### **4.8 Buck-passing and why we ought to be moral**

Since on Broome's theory of pro tanto reasons it must be possible for any pro tanto reason to be outweighed, if moral considerations necessarily determine what we ought to do – i.e. if it is impossible for the normative force of moral considerations to be outweighed – then it follows on Broome's account that moral considerations do not – indeed cannot – provide us with pro tanto reasons.

Consequently, if we accept Broome's theory of pro tanto reasons and if we want to retain our intuition that moral considerations

necessarily override other practical considerations, we are committed to the thesis that the normative authority of moral considerations (i.e. the moral 'ought') cannot be explained by an appeal to pro tanto reasons – moral or otherwise (although I shall have to qualify this claim later). This commitment in turn forces us to look for a non-weighting explanation of why moral considerations are 'ought-making' (if indeed they are). So what is this explanation then? To answer this question, let us return to thick ethical concepts and see if we can, by engaging in a bit of conceptual analysis, find some support for the thesis under consideration.

Providing a conceptual analysis of a concept is generally quite tricky, and providing such an analysis for thick ethical concepts is no exception. In what follows I shall only provide a very rough sketch of what I take some of the ingredient meanings of these concepts to be. My hope is to be able to pin down those features of thick ethical concepts that are relevant to the buck-passing account of wrongness. Ideally, in order to answer the moral question, we should provide such a story for each thick concept, but for obvious reasons, I cannot hope to do so in what remains of this dissertation. Instead, my strategy will be to try to find that feature of 'cruel' which not only explains why actions correctly described as such ought not be done, but which is also shared by all other relevant thick ethical concepts. Before I can do this, however, I need to make some preliminary points.

First (and this goes back to some points made in chapter one), although I have rather sloppily talked about acts and actions as if they were the same thing, thick ethical concepts apply only to actions. In everyday talk, the words 'act' and 'action' are commonly used as if they were interchangeable. In recent philosophical debates, however, these terms have come to denote different things. In order to characterise an event as an action we must specify the reason for which it was done; i.e. we must include the motive, intention, or purpose behind the agent's doing what he did in the description of the action. Thus, as Kant would put it, actions are specified and individuated by their *maxims*: A  $\phi$ -ed *in order to*, or *for the sake of*,  $\psi$ . This means that A's  $\phi$ -ing in order to  $\psi$  is a different action from A's  $\phi$ -ing in order to  $\chi$ . This should not (I hope) be controversial. Suppose A and B are both running through the park. This would be one description of 'what they are doing' (i.e. running through the park), but it would only be a description of what we might call their acts.<sup>53</sup> Very roughly, an act is that which we can, or may, warrantably ascribe to an agent solely on the basis of his overt bodily movements without knowledge of his intentions. This should be contrasted with A's running through the park *in order to* get fit and B's running through the park *in order to* escape from the police. Here we can correctly describe A's and B's *actions* as, respectively, exercising and fleeing because we know the intentions, purposes, or motives behind their running through the park. This

general structure also applies to thick ethical concepts and the actions (as opposed to, merely, acts) they apply to. In what follows I shall nonetheless use these terms interchangeably, meaning by both of them what 'action' means in its distinctively philosophical use.

Second, it may be that the meaning of 'cruel' and other thick concepts (like the meaning of 'game' perhaps) cannot be understood except in a Wittgensteinian 'family resemblance' fashion according to which there is no determinate set of necessary and jointly sufficient conditions that an act has to meet in order to be correctly described as cruel. Nonetheless, we can still specify some of the essential semantic ingredients of 'cruel' without thereby giving a complete analysis of the meaning of the term. In what follows, then, I shall not be trying to provide a full-blown definition of 'cruel'; rather, I shall try to highlight those features of the concept that are relevant to the buck-passing account of wrongness.

To begin with, cruel acts involve intentionally aiming (successfully or unsuccessfully) to harm someone in some way or other.<sup>54</sup> But harming or intending to harm someone cannot be the distinguishing mark of cruel acts as such. Sometimes we cause harm to others unintentionally, and at other times we cause others harm with good intentions and for good reasons, e.g. for the sake of benefiting them in some way or other. In none of these cases

would we describe such acts as cruel. (Incidentally, then, it is simply false that 'one sometimes has to be cruel to be kind'.) For an action to be cruel, the harm caused or intended (or both) by the action must be of a particular kind. My proposal is that 'A's  $\phi$ -ing was cruel' means, at least in part,

A intentionally aimed (successfully or unsuccessfully) to inflict harm by  $\phi$ -ing for some reason, and A ought not have  $\phi$ -ed to the extent that, or in the manner that, or for the reason that he did.

The main idea is that we cannot possibly hope to characterise a cruel action without essential reference to the idea that the action in question ought not have been pursued or ought not have been pursued in the way it was.

Consider another example: 'unfair'. John Broome has suggested, plausibly, that fairness requires that people's claims (to a particular good) 'should [*ought to*] be satisfied in proportion to their strength.'<sup>55</sup> This characterisation of fairness plausibly implies that if an act is unfair, then people's claims have not been satisfied in the way they ought to have been. I'm not sure whether Broome would be willing to say that this is a conceptual truth, but the point here is simply that the concept of fairness somehow involves or incorporates the concept 'ought'. To say of A's action that it was



unfair is to say, roughly, that A treated someone's legitimate claim to a particular good in a way that, or for a reason that, he ought not have. Here it also seems appropriate to add that if Rawls is correct (which he undoubtedly is) in saying that justice is, essentially, a matter of fairness, then considerations of justice are plausibly ought-making as well. Again, I think 'A acted unjustly' means, in part, 'A ought not have acted in the way he did'. To mention a final example: 'murder'.

Murdering someone does not only involve killing, it involves killing someone in a particular way, or more accurately perhaps, for some unspecified *illegitimate* reason (or motive). Surely it is no coincidence that someone who (in a court of law) has been proven to have killed in self-defence or in order to protect the life of others will not be found guilty of murder. It is only if the killing was prompted by some motive which under the circumstances *ought not* have prompted the killing that we can correctly describe the case as one involving murder. This does not mean of course that there cannot be reasonable disagreements about whether a particular event should be classified as murder or not; but this in turn should be taken to show that there can be reasonable disagreements about whether an agent's having acted in a certain way really was wrong. This being so, however, should not distract us from the fact that there are perfectly clear instances of murder – and hence of wrong-

doing. The same points apply, I believe, to the other concepts mentioned above.

These analyses are of course not very precise and more would have to be said in order to fully vindicate the buck-passing account of wrongness. Nonetheless, I dare say that when someone acts wrongly this person is in some way or other acting either from a motive which he ought not act on or in a manner in which he ought not be acting. So a person who acts wrongly is never fully practically justified in so acting. In one sense, then, Scanlon is absolutely right: an act is wrong if and only if it cannot be justified to others (or himself for that matter). But the connection between wrongness and justifiability is, on my account, established by the buck-passing account of wrongness and a conceptual analysis of the relevant thick ethical concepts. Such a conceptual analysis should reveal that the relevant thick concepts incorporate the property 'ought-not-be-performed'. In essence, then, the explanation of why we ought not do those things that can be accurately described as cruel, unfair, unjust, inconsiderate etc. is that it is true in virtue of the meaning of these concepts that actions accurately described by them ought not be performed in the manner, or to the extent, or for the reason which they were performed.

So what about the ought fact 'we ought to be moral' - what explains it? The correct answer is that the explanation of why we ought to be moral is that being moral just is being sensitive to the

fact that certain properties of various actions are genuinely ought-making. But why ought we be sensitive to, and act consistently with the recognition that, certain considerations are genuinely ought-making? Well, simply because these considerations are genuinely ought-making, that's why! By acting in accordance with the true normative beliefs we acquire by being sensitive to moral considerations, we are fully practically justified in acting morally.<sup>56</sup> There is no appeal to any *pro tanto* reasons in this explanation of why we ought to be moral. We ought to be moral because being moral consists in doing what we ought to do, and for the right reasons.

#### **4.9 Problems and prospects for the buck-passing account**

I suspect that much of the scepticism about morality – about, among other things, its metaphysical, normative, and epistemic status – stems from a scepticism about the supposedly queer property of wrongness. I think that many of the metaphysical worries we might have about morality and wrongness can be eased if we accept the buck-passing account of wrongness. Whatever metaphysical scepticism may remain after the buck-passing account has been accepted, much if not all of it can be allayed by, as I said in chapter one, rejecting the correspondence theory of truth in favour of minimalism (or *disciplined syntacticism* as it is sometimes

called). In this chapter, and in the last few sections especially, I have provided some arguments that hopefully show that moral considerations are not only genuinely normative but overridingly so. I need to add quite an important qualification to this claim, but before I do that, I want to say something about the epistemological scepticism about morality and wrongness. To address this worry, consider again Thomson's challenge to Scanlon's contractualism:

I cannot bring myself to believe that what *makes* it wrong to torture babies to death for fun (for example) is that doing this 'would be disallowed by any system of rules as a basis for the general regulation of behavior which no one could reasonably reject as a basis for informed, unforced general agreement.' My impression is that the explanation goes in the opposite direction - that it is the patent wrongfulness of the conduct that explains why there would be general agreement to disallow it.

My contention, of course, is that torturing babies for fun is wrong because doing so is cruel, and the fact that torturing babies for fun is cruel is also what explains why we could not reasonably reject principles disallowing such acts. If Thomson wants to insist that it is the 'patent wrongfulness' of the act which explains why doing this thing is not justifiable, then absent some story about what makes

an action wrong, the familiar charge that wrongness is epistemically queer still retains its force. Were she to accept the buck-passing account of wrongness, however, she could still reject the idea that torturing babies for fun is wrong because doing so would be disallowed by principles we could not reasonably reject. The added benefit of doing so is that she would also be able to tell a more plausible epistemological story about how we come to have knowledge about the wrongness of such acts.

Although this needs to be worked out in more detail, it seems to me that whatever story we are prepared to offer as an account of how we come to know that a particular, relevant, thick ethical concept applies to a particular action, the same story can (in conjunction with the buck-passing account of wrongness) be told about how we can come to know that an act is wrong. There may actually be reason to feel optimistic about the prospects of telling such a story. Consider the following passage from Williams:

[A thick ethical concept] may be rightly and wrongly applied, and people who acquired it can agree that it applies or fails to apply to some new situation. In many cases the agreement will be spontaneous, while in other cases there is room for judgement and comparison. Some disagreement at the margin may be irresolvable, but this does not mean that the use of the concept is not controlled

by the facts or by the users' perception of the world. (As with other concepts that are not totally precise, marginal disagreements can indeed help to show how their use *is* controlled by the facts.)<sup>57</sup>

This of course fits nicely with the points made earlier about murder etc. If my arguments in this dissertation are correct – especially if, as I believe, it is true that propositions like 'you ought not act cruelly' are conceptual truths – whatever epistemic scepticism about morality remains can only sensibly be viewed as a scepticism about whether some actions really are cruel, unjust, unfair, inconsiderate etc. But if these concepts in turn are, as Williams suggests, sufficiently world-guided to ensure general agreement on at least their 'non-marginal' application, many of the epistemic worries that have traditionally haunted moral philosophers can perhaps finally be laid to rest.

I turn now to the qualification of the buck-passing account I promised to say something about. It may be objected that the buck-passing account may lead to some unacceptable and overly moralistic conclusions. Although I have said that moral considerations necessarily outweigh other considerations, this is unfortunately not strictly true. The fact that I have promised something and the fact that I would be lying if I said that I hope Denmark do well in Euro 2004 are certainly moral considerations,

but the concepts 'promise' and 'lie' do not seem to have, as part of their meaning, the idea that we ought not do these things. To take a familiar example, suppose A has promised B to meet her at the cinema at certain time. On the way to the cinema A comes across a terrible traffic accident and he is the first one on the scene. If he were to stop and help the victims he would not be able to keep his promise to B. If promises are ought-making, A ought not stop and help the traffic accident victims. Yet this is absurd, clearly we ought to stop and the help the traffic accident victims. Therefore promises cannot be ought-making and if the buck-passing account implies this it should be rejected.

The correct response to this challenge is to admit that not all properties accurately expressed by some thick ethical concept are ought-making. But the buck-passing account of wrongness is consistent with the claim that some properties accurately referred to by thick ethical concepts may only be pro tanto reason-giving. But this, in turn, should be taken to show that those actions to which thick ethical concepts of (merely) the pro tanto reason-giving type correctly apply are not going to be wrong (if they are) in virtue of having *these* properties. We should therefore draw a distinction between what we (somewhat inelegantly) might call strong and weak thick ethical concepts. Strong thick ethical concepts are those I have already discussed: cruel, unfair, unjust etc., concepts which have the idea of 'ought-not be doneness' built into them. Weak



thick ethical concepts, on the other hand, concepts such as 'lying' and 'promising', are only pro tanto-reason giving since these concepts do not seem to have, as part of their meaning, that idea that we ought not do these things. Alternatively, we could say that promises are ought-making, but only defeasibly so (this, of course, means that they are pro tanto-reason giving). A good case can be made for saying that promises have a defeasibility condition built into them. After all, this is how promises work in everyday life. When A promised B that he would meet her at the cinema, it would be absurd to think that A has thereby promised to meet B at the cinema *no matter what*. A has not made such a promise; rather what he has promised (at least tacitly) is that he will meet B at the cinema *unless* certain conditions obtain. This defeasibility condition is, and perhaps must be, understood (again, at least tacitly) by all promisers and promisees - it is part of the institution of promising. So the obligation to keep one's promises is conditional, and the defeasibility conditions of a given promise may depend on the importance of the promise being kept and the circumstances in which it was made etc. It should also be pointed out that the defeasibility conditions cannot plausibly be thought of only as teleological in nature. If A has promised B to meet her at the cinema but then remembers that he has made an earlier promise to C to meet her at the pub, then it may be that A should keep the promise he made first simply because he made the promise to C

first. It is of course true that in this case A incurs an additional obligation to apologise (or at least explain his actions) to B. But it would be false to say that promises can be defeated only by considerations relating to the goodness of the consequences of breaking a promise. I think a similar story can be told about lying.

Buck-passers should not be afraid to admit that the justifiability of an act may sometimes depend on the value of the outcomes of that action and sometimes it may not. Whether or not an agent is able to correctly determine whether an action would be wrong will to a large extent depend on that agent's capacity for correctly determining the 'salience and shape' (to borrow some terminology from Dancy) of the normative and non-normative features of the situation in which we find ourselves. Perhaps buck-passers should say that you ought not act in such a way that your action could accurately be described by using a weak thick ethical concept unless your failure or refusal to act in that way could accurately be described by using a strong thick ethical concept.<sup>58</sup> This should be taken merely as a rule of thumb however. As far as the buck-passing account of wrongness is concerned, it may even be morally permissible to break a promise or tell a lie if there were sufficiently strong prudential reasons to do so. Nonetheless, even if we have strong prudential reason to  $\phi$ , if  $\phi$ -ing can accurately be described by using a strong thick ethical concept, we ought not do it.<sup>59</sup>

It may be objected at this point that I have provided no general method for distinguishing weak and strong ethical concepts. I accept this objection. At present I know of no principled way of distinguishing the two kinds of concepts. This is something which remains for the buck-passer to work out. It may of course be the case that there simply is no general way of distinguishing the two; it may turn out that there are no other distinguishing feature of these concepts besides their being ought-making and pro tanto-reason giving, respectively. Perhaps the best we can do is simply to look at them one concept at a time and see how they are used in ordinary language. Since my aim in this chapter has been merely to provide a sketch of what a successful buck-passing account of wrongness will look like, I am not too worried about this – although I do of course acknowledge the importance of the objection.

Finally, I want to briefly say something about two rival accounts of wrongness. I said earlier that a generic version of the buck-passing account of wrongness should be acceptable to a wide range of moral theorists. I also said that these theorists will disagree about what the relevant reason-giving lower-order properties of wrong actions are. Utilitarians will say that there is only one relevant lower-order property: the property of failing to maximise utility (whatever that happens to be). Utilitarians will therefore think that the account I have given is wrong. To this I can only reply that the most commonly cited, and to my mind

correct, reason for rejecting utilitarianism is precisely that it fails to acknowledge and give due normative weight to those properties of certain actions that be accurately described by, or referred to by, thick ethical concepts. To repeat a line from Prichard: 'the balance of resulting good may be, and often is, not on the side of justice.'<sup>60</sup> For a utilitarian, the fact that an action is cruel, unfair, unjust, inconsiderate, or selfish has little or no significance when it comes to determining what we ought to do. This is exactly why utilitarianism strikes many, myself included, as implausible.

My account is compatible with many things Kantians want to say. In the Kantian tradition, autonomy plays a central role. It may well be that, properly understood, this concept is also both 'world-guided' and 'action-guiding'. To say of an action that it would violate someone's autonomy (by treating him as a mere means perhaps; and thus, for Kantians, to treat him wrongly) could plausibly be interpreted along buck-passing lines as meaning that it would be inconsiderate, unfair, or cruel or whatever the appropriate description might be. If any of these concepts apply to actions in which we treat others as mere means, this would mean that we could not justifiably treat this person in this way (whatever way that is). It may well turn out that those who accept my account of wrongness could agree with most sensible Kantians about the extension of the predicate 'is wrong'. More would have to be said in order to show that this is so, but the point I am trying to highlight

here is that my account of wrongness and the substantive moral principles that follow from it will not, I suspect, diverge too much from what the Kantian thinks is right. If not entirely Kantian, my account is at least decidedly non-consequentialist.

Lastly, let me briefly say something about the buck-passing account of wrongness and how it relates to moral phenomenology. I think Scanlon's appeal to the Kantian-flavoured ideal of standing in a justifiability relation to others (or, as Kant would have put it, to be related to others by way of being a legislative member of a Kingdom of Ends) is absolutely correct in this context. More precisely, I think it is our deep desire to live up to this ideal (and hence to act only in ways that could be justified to others) which to a large extent explains the particular phenomenology that accompanies judgements of right and wrong. When we judge that some act would be wrong (or when we realise that we have acted wrongly) this judgement, or this recognition, is – at least for most of us, and for most of the time – accompanied by the feeling that in so acting we would in some way or other alienate ourselves from not only the rest of the moral community but from ourselves as well. If we cannot justify our acting in a certain way to others, then we cannot justify our acting in that way to ourselves either. This self-alienating aspect of morality is brought out clearly by the common thought that 'I couldn't live with myself if I did that'.

The essentially *moral* feelings of guilt, blame, and remorse etc dispose, as John Skorupski puts it, 'to withdrawal of recognition, casting out of the community'.<sup>61</sup> This moral community (or Kingdom of Ends), in turn, is the one to whose members most of us, most of the time, strive to live in justificatory harmony with. This, in turn, can be done only if we refrain from doing that which can accurately be described as cruel, or unfair, or unjust – and, hence, wrong.

---

<sup>1</sup>Hurley, pp. 151-2.

<sup>2</sup>Hooker, p. 43.

<sup>3</sup>Scanlon (1998), p. 96. My italics. Although Scanlon's buck passing account of value has generated a great deal of controversy, he is not the originator of this kind of value analysis. Various versions of the buck passing account can be found in Brentano, Broad, Ross, Ewing, Rawls, Chisholm, Wiggins, and Lemos. For a brief historical exposition of these authors' views, see Rabinowicz and Rønnow-Rasmussen.

<sup>4</sup>Scanlon (1998), p. 99.

<sup>5</sup>*Ibid.*, p. 97. My italics.

<sup>6</sup>'Naturalistic' does not in the present context mean 'empirical'. 'Naturalism' should here be understood as an umbrella doctrine which encompasses all theories that reject the existence of *sui generis* 'value facts'. According to this definition of 'naturalism', Divine Command Theory, for example, would be a naturalistic value (or moral) theory. 'Naturalism' is sometimes (and perhaps more frequently these days) used in a narrower sense. In this sense, 'naturalism' is the ontological doctrine that only natural properties exists, and that natural properties are all and only those properties countenanced by natural science. See Pigden.

<sup>7</sup>Here I intend 'natural' to mean 'countenanced by natural science'.

<sup>8</sup>See Rabinowicz and Rønnow-Rasmussen. See also Parfit (2002).

<sup>9</sup>Scanlon (1998), p. 96.

<sup>10</sup>Dancy has recently argued that we cannot analyse value in terms of reasons. See Dancy (2000b).

<sup>11</sup>Scanlon himself refers to this problem as 'Prichard's dilemma', but I think the dilemma Scanlon discusses is not the same as the one Prichard thought he had identified.

<sup>12</sup>*Ibid.*, pp. 149-50.

<sup>13</sup>In Scanlon (1996) Scanlon (following Kurt Baier) suggests that the question 'why be moral?' might be asked as a request for an elucidation of the relation between moral and nonmoral reasons. I will not be discussing this interpretation here.

<sup>14</sup>Scanlon (1998), p. 147.

<sup>15</sup>*Ibid.*, pp. 147-8.

<sup>16</sup>*Ibid.*, p. 148.

<sup>17</sup>*Ibid.*

<sup>18</sup>*Ibid.*

<sup>19</sup>*Ibid.*, p. 149.

<sup>20</sup>This formula comes from Scanlon's characterisation of contractualism on p. 153 of Scanlon (1998). Here Scanlon presents the formula in such a way that the justifiability condition provides a



sufficient, but not a necessary condition of an act's being wrong. The reason why Scanlon does not present his contractualist formula as a biconditional in this context is that he is (here) only interested in wrongness in the sense of 'what we owe to each other' - i.e. what we owe to other *agents*. There may be other kinds of wrongness that are not captured by Scanlon's formula - the wrongness resulting from cruelty toward animals for instance. If the justifiability condition also specified a necessary condition for an act's being wrong, Scanlon would not be able to explain why for example torturing animals for fun is wrong. For a discussion of this topic see Scanlon (1998), pp. 177-88. Also, and importantly, the clause I have italicised does not appear in the particular formulation of contractualism I have quoted. However, Scanlon makes it clear in other passages of his book that this clause is vital to his contractualism. It is particularly important for his discussion of aggregation. See Scanlon (1998), pp. 229ff.

<sup>21</sup>In what follows I have been helped greatly by reading the following papers: Stratton-Lake (2003), Stratton-Lake (2003b), Ridge (2001), Ridge (2003), McNaughton and Rawling (2003).

<sup>22</sup>*Ibid.*, p. 170. My italics.

<sup>23</sup>Thomson, p. 30, n. 19. See also Pettit (1999), pp. 7-8, and McGinn (1999), pp. 34-8.

<sup>24</sup>Scanlon (1998), p. 11.

<sup>25</sup>Ridge, pp. 472-73.

<sup>26</sup>Scanlon (1998), p. 391, n. 21.

<sup>27</sup>*Ibid.*, p. 11. My italics.

<sup>28</sup>See Stratton-Lake (2003).

<sup>29</sup>Scanlon (1998), p. 11.

<sup>30</sup>Interestingly, there are strong similarities between the buck-passing account of wrongness (and value, perhaps) and 'program explanations' (as developed by Frank Jackson and Philip Pettit) of causally relevant (as opposed to causally efficacious) properties in natural science. See Jackson and Pettit.

<sup>31</sup>Williams (1985), p. 140.

<sup>32</sup>Williams (1995), pp. 37-8.

<sup>33</sup>See Scanlon (1998), p. 187.

<sup>34</sup>*Ibid.*, p. 161. What Scanlon (somewhat misleadingly) here calls the problem of moral motivation is what I take to be the problem posed by the moral question.

<sup>35</sup>*Ibid.*, p. 162. Given that Scanlon (rightly) accepts the buck-passing account of value (as an account of the normatively epiphenomenal character of value), it is odd (to say the least) that he goes on to claim that it is the (very great?) *value* of standing in a justifiability relation to others that gives us overriding reason to be related to others in this way. To be consistent, Scanlon should say that it is the properties in virtue of which it is valuable to stand

in this relation to others that give us reason to be related to others in this way. Scanlon does not, to my knowledge anyway, say anything about what these properties might be.

<sup>36</sup>One could easily imagine a friendship between two evil persons in which the 'dictates of friendship' were not reason-giving *at all*.

<sup>37</sup>If these reasons are genuine reasons, however, we are still owed some account about why this is so - after all, it is precisely the claim that moral considerations have reason-giving force that the amoralist denies (even though Scanlon is not interested in the amoralist's challenge).

<sup>38</sup>Stratton-Lake disagrees. He says: 'Scanlon's contractualist principle goes beyond the mere idea that we have moral reasons, and aspires to tell us the general nature of those reasons. Moreover, it aims to characterize their nature in such a way as to connect the importance of moral reasons with the value of an ideal of mutual recognition via the notion of justifiability to others. [...] It characterizes...moral reasons...in such a way as to make possible a Kantian explanation of their importance and priority.' (2003b, pp 336-7) I think the appeal to the ideal of mutual recognition is somewhat misplaced. The ideal of standing in a justifiability relation to others adds nothing significant to the reasons we have for being moral; rather, our standing in this relation to others is, as I have already said, a consequence of acting morally. However, I *do* think this ideal, or more accurately, our deep desire to live up to this ideal, can explain the peculiar phenomenology that attends moral judgements. I will return briefly to this idea toward the end of this chapter.

<sup>39</sup>For an important discussion of (moral) fetishism and judgement internalism and externalism, see Smith (1994), pp. 71-6.

<sup>40</sup>Scanlon (1998), p. 152.

<sup>41</sup>Williams (1976), p. 18. I have here taken Williams' slogan slightly out of the context in which he employs it.

<sup>42</sup>Raz (1999), p. 67.

<sup>43</sup>Broome (forthcoming), p. 1.

<sup>44</sup>*Ibid.*

<sup>45</sup>*Ibid.*, p. 3.

<sup>46</sup>To call these kinds of entities 'ought facts' does not entirely square with what I said in chapter one about the metanormative assumptions I am working with in this dissertation. Given the assumptions I made in chapter one, it would be more correct (or consistent) to call these things 'ought truths' rather than 'ought facts', but for simplicity, and to avoid unnecessary confusion, I shall simply adopt Broome's terminology.

<sup>47</sup>As far as I'm aware, the concept of a perfect reason is an invention of Broome's; it is certainly not commonly used in the literature.

<sup>48</sup>*Op. Cit.*, p. 8.

<sup>49</sup>*Ibid.*, p. 9.

<sup>50</sup>Scanlon (1998), p. 17.

<sup>51</sup>*Op. Cit.* See pp. 2-3.

<sup>52</sup>*Ibid.*, pp. 11-12.

<sup>53</sup>For more on this distinction and its significance, see Korsgaard (2002), Lecture One.

<sup>54</sup>I should emphasize that a cruel act need not actually cause any harm. If, for instance, I pour poison into the city's water supply with the intention of increasing my sales of antidotes, then this act counts as cruel even if some hero manages to neutralise the danger before anyone is (actually) harmed.

<sup>55</sup>Broome (1999b), p. 116.

<sup>56</sup>Or, at least, very nearly fully practically justified. In order to be fully practically justified it may not be enough that we act consistently with our true (strongest) normative judgement. It may well be that we have to include some *epistemic* 'truth-tracking' constraint on the acquisition of our normative beliefs in order for our action to be fully practically justified. I shall not address this concern here.

<sup>57</sup>Williams (1985), pp. 140-1.

<sup>58</sup>Sometimes it may be cruel to tell the truth ('does my bum look big in this?'), and in such cases there can be nothing wrong with telling a (white) lie, in fact, given certain circumstances it may even be required.

<sup>59</sup>Alternatively, we could say that on any plausible theory of prudence we could never have sufficient prudential reason to act cruelly, unfairly, and unjustly etc. The idea here would be that an agent's life could not plausibly be considered to 'go better' if he were to do these things. Some philosophers may reject such a theory of prudence and prudential reasons on the grounds that it would be too moralistic. I disagree with this. I think such a theory has a lot going for it.

<sup>60</sup>Prichard, p. 5.

<sup>61</sup>Skorupski (1999), p. 180.

## BIBLIOGRAPHY

- Altham, J. E. J. (1986): 'The legacy of Emotivism' in Graham McDonald and Crispin Wright (eds).
- Altham J. E. J., and Harrison, Ross (eds) (1995): *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams* (Cambridge: Cambridge University Press).
- Audi, Robert (1998): *Epistemology; An Introduction to Contemporary Theory of Knowledge* (London: Routledge).
- Bermúdez, José, and Millar, Alan (eds) (2001): *Reason and Nature: Essays in the Theory of Rationality* (Oxford: Oxford University Press).
- Blackburn, Simon (1988): 'Attitudes and contents', *Ethics*, 98.
- Blackburn, Simon (1992): 'Through Thick and Thin', *Proceedings of the Aristotelian Society*, Supplementary Volume 61.
- Blackburn, Simon (1995): 'Practical Tortoise Raising', *Mind*, 104.
- Bogdan, R. J. (ed.) (1986): *Belief* (Oxford: Oxford University Press).
- Bratman, Michael (1987): *Intentions, Plans, and Practical Reason* (Cambridge, Mass.: Harvard University Press).
- Broome, John (1997): 'Reasons and Motivation', *Proceedings of the Aristotelian Society*, Supplementary Volume 71.
- Broome, John (1999): 'Normative Requirements', *Ratio*, Vol. XII.

- Broome, John (1999b): 'Fairness', reprinted in his *Ethics out of Economics* (Cambridge: Cambridge University Press, 1999).
- Broome, John (2001): 'Practical Reasoning' in José Bermúdez and Alan Millar (eds).
- Broome, John (2001b): 'Normative Practical Reasoning', *Proceedings of the Aristotelian Society*, Supplementary Volume 75.
- Broome, John (forthcoming): 'Reasons' in R. Jay Wallace, Michael Smith, Samuel Scheffler, and Philip Pettit (eds).
- Broome, John (unpublished): 'Reasoning'.
- Brink, David (1997): 'Kantian Rationalism' in Garrett Cullity and Berys Gaut (eds).
- Carroll, Lewis (1895): 'What the Tortoise said to Achilles', *Mind*, 4.
- Cullity, Garrett, and Gaut, Berys (eds) (1997): *Ethics and Practical Reason* (Oxford: Clarendon Press).
- Dancy, Jonathan (1993): *Moral Reasons* (Oxford: Blackwell).
- Dancy, Jonathan (1995): 'In Defence of Thick Concepts', in Peter French (ed.).
- Dancy, Jonathan (2000): *Practical Reality* (Oxford: Oxford University Press).
- Dancy, Jonathan (2000b): 'Should we pass the buck?' in O'Hear (ed.).
- Darwall, Stephen (1990): 'Autonomist Internalism and the Justification of Morals', *Nous*, 24.

- Darwall, Stephen (1997): *Reason, Motives, and the Demands of Morality: An Introduction* in Stephen Darwall, Alan Gibbard, and Peter Railton (eds).
- Darwall, Stephen; Gibbard, Alan; and Railton, Peter (eds) (1997): *Moral Discourse and Practice* (Oxford: Oxford University Press).
- Davidson, Donald (1994): Entry in Samuel Guttenplan (ed.).
- Davidson, Donald (2001): *Essays on Actions and Events* (Oxford: Oxford University Press).
- Dreier, James (1997): 'Humean Doubts about the Practical Justification of Morality' in Garrett Cullity and Berys Gaut (eds.).
- Egonsson, Dan; Josefsson, Jonas; Petersson, Björn; and Rønnow-Rasmussen, Toni (eds) (2002): *Exploring Practical Philosophy: From Action to Values* (Ashgate).
- Field, Hartry (1986): 'Tacit Beliefs' in R. J. Bogdan (ed.).
- Frankena, W. K. (1963): *Ethics* (Englewood Cliffs, New Jersey: Prentice Hall).
- French, Peter (ed.) (1995): *Moral Concepts* (*Midwest Studies in Philosophy*, 20).
- Gauthier, David (1986): *Morals by Agreement* (Oxford: Clarendon Press).
- Geach, Peter (1965): 'Assertion', *Philosophical Review*, 74.



- Gibbard, Allan (1992): 'Thick Concepts and Warrant for Feelings', *Proceedings of the Aristotelian Society*, Supplementary Volume 66.
- Goldman, A. I. (1967): 'A causal theory of knowing', *Journal of Philosophy*, 64.
- Guttenplan, Samuel (ed.) (1994): *A Companion to the Philosophy of Mind* (Oxford: Blackwell's).
- Guyer, Paul (ed.) (1992): *The Cambridge Companion to Kant* (Cambridge: Cambridge University Press).
- Haldane, John, and Wright, Crispin (eds) (1993): *Reality, Representation, and Projection* (New York: Oxford University Press).
- Hale, Bob: 'Can there be a logic of attitudes?' in John Haldane and Crispin Wright (eds).
- Hooker, Brad (1987): 'Williams' Argument Against External Reasons', *Analysis*, 47.
- Hume, David (1975): *An Enquiry Concerning the Principles of Morals*, edited by L. A. Selby-Bigge and P. H. Nidditch, 3rd edn (Oxford: Clarendon Press).
- Hume, David (1978): *A Treatise of Human Nature*, edited by L. A. Selby-Bigge and P. H. Nidditch, 2nd edn (Oxford: Clarendon Press).
- Hurley, Susan (2001): 'Reasons and motivation: the wrong distinction?', *Analysis*, 61.



- Jackson, Frank (1998): *From Metaphysics to Ethics* (Oxford: Clarendon Press).
- Jackson, Frank, and Pettit, Philip (1990): Program explanation: a general perspective, *Analysis*, 50, vol. 2.
- Kant, Immanuel (1948): *Groundwork of the Metaphysics of Morals*, translated by H. J. Paton (London: Routledge).
- Korsgaard, Christine (1996): *The Sources of Normativity* (Cambridge: Cambridge University Press).
- Korsgaard, Christine (2002): *Self-Constitution: Action, Identity, and Integrity* (The Locke Lectures, University of Oxford).
- Lycan, William G. (1986): 'Tacit Belief' in R. J. Bogdan (ed.).
- Mackie, John (1977): *Ethics: Inventing Right and Wrong* (Harmondsworth: Penguin).
- McDonald, Graham, and Wright, Crispin (eds) (1986): *Fact, Since, and Morality: Essays on A. J. Ayer's Language, Truth and Logic* (Oxford: Blackwell's).
- McDowell, John (1978): 'Are Moral Requirements Hypothetical Imperatives?', *Proceedings of the Aristotelian Society*, Supplementary Volume 52.
- McDowell, John (1995): 'Might there be External Reasons?' in J. E. J. Altham and Ross Harrison (eds).
- McGinn, Colin (1999): 'Reasons and unreasons', *The New Republic*, 24 May.
- McNaughton, David (1988): *Moral Vision* (Oxford: Blackwell's).

- McNaughton, David, and Rawling, Piers (2003): 'Can Scanlon avoid redundancy by passing the buck?', *Analysis*, 63, vol. 4.
- Miller, Alexander (1998): *Philosophy of Language* (London: UCL Press).
- Millgram, Elijah (ed.) (2001): *Varieties of Practical Reasoning* (Cambridge, Mass: MIT Press).
- Nagel, Thomas (1970): *The Possibility of Altruism* (Princeton: Princeton University Press).
- Nozick, Robert (1981): *Philosophical Explanations* (Cambridge, Mass: Harvard University Press).
- O'Hear, Anthony (ed.) (2000): *Philosophy, the Good, the True and the Beautiful* (Cambridge: Cambridge University Press).
- Parfit, Derek (1997): 'Reasons and Motivation', *Proceedings of the Aristotelian Society*, Supplementary Volume 71.
- Parfit, Derek (2001): 'Rationality and Reasons' in Dan Egonsson, Jonas Josefsson, Björn Petersson, and Toni Rønnow-Rasmussen (eds).
- Parfit, Derek (forthcoming): *Rediscovering Reasons*.
- Parfit, Derek (unpublished): 'Kant's Formulae'.
- Pettit, Philip (1999): 'Doing unto others', *Times Literary Supplement*, 25 June.
- Pigden, Charles: 'Naturalism' in Singer (ed.).
- Price, H. H. (1947): 'Harold Arthur Prichard, 1871-1947', *Proceedings of the British Academy*, 33.

- Prichard, H. A. (1947): 'Does Moral Philosophy Rest on a Mistake?' in Ross (ed.).
- Rabinowicz, Wlodek, and Rønnow-Rasmussen, Toni (unpublished): 'Fitting Pro Attitudes and Value'.
- Railton, Peter (1996): 'Moral Realism: Prospects and Problems' in Sinnott-Armstrong (ed.).
- Rawls, John (1993): *Political Liberalism* (New York: Columbia University Press).
- Raz, Joseph (1986): *The Morality of Freedom* (Oxford: Oxford University Press).
- Raz, Joseph (1999): *Engaging Reason: On the Theory of Value and Action* (Oxford: Oxford University Press).
- Ridge, Michael (2001): 'Saving Scanlon: contractualism and agent-relativity', *The Journal of Political Philosophy*, 9.
- Ridge, Michael (2003): 'Contractualism and the new and improved redundancy objection', *Analysis*, 63, vol. 4.
- Ross, W. D. (ed.) (1949): *Moral Obligation, Essays and Lectures* (Oxford: Clarendon Press).
- Scanlon, Thomas (1996): 'Self-anchored Reasons' in J. B. Schneewind (ed.).
- Scanlon, Thomas (1998): *What We Owe to Each Other* (Cambridge, Mass: Harvard University Press).

- Schneewind, J. B. (ed.) (1996): *Reason, Ethics, and Society: Themes from Kurt Baier with his Responses* (Chicago: Open Court).
- Segal, Gabriel (1994): 'Epistemology of Belief (2)' in Samuel Guttenplan (ed.).
- Sen, Amartya (2000): 'Consequential Evaluation and Practical Reason', *Journal of Philosophy*, 9.
- Singer, Peter (ed.) (1991): *A Companion to Ethics* (Oxford: Blackwell's).
- Sinnott-Armstrong, Walter (ed.) (1996): *Moral Knowledge?* (New York: Oxford University Press).
- Skorupski, John (1999): 'Irrealist Cognitivism', *Ratio*, Vol. XII.
- Skorupski, John (1999b): 'Freedom, Morality, and Recognition', in his *Ethical Explorations* (Oxford: Oxford University Press).
- Skorupski, John (forthcoming): 'Internal Reasons and the Scope of Blame', in Thomas (ed.).
- Skorupski, John (unpublished): 'Irrealism about reasons'.
- Smith, Michael (1994): *The Moral Problem* (Oxford: Blackwell's).
- Sosa, Ernest, and Kim, Jaegwon (eds) (2000): *Epistemology: An Anthology* (Oxford: Blackwell's).
- Stratton-Lake, Philip (2003): 'Scanlon's contractualism and the redundancy objection', *Analysis*, 63, vol. 1.

- Stratton-Lake, Philip (2003b): 'Scanlon, permissions, and redundancy: response to McNaughton and Rawling', *Analysis*, 63, vol. 4.
- Thomas, Alan (ed.) (forthcoming): *The Cambridge Companion to the Philosophy of Bernard Williams* (Cambridge University Press).
- Thomson, Judith (1990): *The Realm of Rights* (Cambridge, Mass: Harvard University Press).
- Vallentyne, Peter (ed.) (1991): *Contractarianism and Rational Choice: Essays on David Gauthier's Morals By Agreement* (Cambridge: Cambridge University Press).
- Van Cleve, James (2000): 'Foundationalism, Epistemic Principles, and the Cartesian Circle' in Ernest Sosa and Jaegwon Kim (eds).
- Velleman, David (2000): *The Possibility of Practical Reason* (Oxford: Oxford University Press).
- Wallace, Jay; Smith, Michael; Scheffler, Samuel; and Pettit, Philip (eds) (forthcoming): *Reason and Value: Essays on the Moral Philosophy of Joseph Raz* (Oxford: Oxford University Press).
- Wiggins, David (1987): 'A Sensible Subjectivism', in his *Needs, Values, and Truth* (Oxford: Blackwell's).
- Williams, Bernard (1976): *Problems of the Self* (Cambridge: Cambridge University Press).

Williams, Bernard (1981): *Moral Luck* (Cambridge: Cambridge University Press).

Williams, Bernard (1985): *Ethics and the Limits of Philosophy* (London: Collins).

Williams, Bernard (1995): *Making Sense of Humanity* (Cambridge: Cambridge University Press).

Williams, Bernard (1995b): 'Replies' in J. E. J. Altham and Ross Harrison (eds).

Williams, Bernard (2001): 'Postscript: Some Further Notes on Internal and External Reasons', in Elijah Millgram (ed.).

Wright, Crispin (1992): *Truth and Objectivity* (Cambridge, Mass: Harvard University Press).